Statistica Sinica **14**(2004), 249-267

CHECKING FOR THE GAMMA FRAILTY DISTRIBUTION UNDER THE MARGINAL PROPORTIONAL HAZARDS FRAILTY MODEL

Sufang Cui and Yanqing Sun

Transamerica Occidental Life Insurance and
University of North Carolina at Charlotte

Abstract: The marginal proportional hazards frailty model for multivariate failure time data characterizes the intracluster dependency with the frailty model while formulating the marginal distributions with the proportional hazards model. The gamma frailty distribution has been widely used to model intracluster dependency because of its simple interpretation and mathematical tractability. Glidden (2000) proposed a two-stage method for estimating the dependence parameter under the marginal proportional hazards frailty model when the frailty follows a gamma distribution. The goodness of fit test for the marginal proportional hazards model has been proposed by Spiekerman and Lin (1996). In this paper, we provide a graphical as well as a numerical method for checking the adequacy of the gamma frailty distribution. The test process is derived from the posterior expectation of the frailty given the observable data. The critical value can be obtained by a Monte Carlo simulation. Two examples from genetics studies are provided to illustrate the proposed testing procedure.

Key words and phrases: Correlated failure times, posterior expectations, tests.

1. Introduction

Multivariate failure time data arise when there are clusters and the subjects within each cluster are correlated. Examples include the tumor occurrence in the litter-matched experiment, the onset of schizophrenia among family members and the development of a disease to paired organs. In many statistical studies, it is of particular interest to investigate the intracluster dependency. There has been a substantial research effort toward describing the dependency structure for multivariate failure time data in recent years. A number of authors, including Vaupel, Manton and Stallard (1979), Clayton and Cuzick (1985), Hougaard (1987), Oakes (1989) and Nielsen, Gill, Andersen and Sørensen (1992), have studied frailty models in which the nature of dependence is modeled parametrically. However, these models pose considerable difficulties with respect to estimation and goodness of fit. To bypass the difficulties caused by the intracluster dependency, Wei, Lin and

Weissfeld (1989) proposed the marginal proportional hazards model. It assumes that the marginal distributions for the correlated failure times satisfy the Cox model while leaving the dependence structure completely unspecified. Spiekerman and Lin (1998) extended this idea to a more general setting and established a large sample theory for the estimates of regression parameters and cumulative baseline hazard functions.

Suppose that there are n independent clusters. Each cluster consists of K distinct failure types, each of which has L exchangeable failure times. Let T_{ikl} and C_{ikl} denote the failure and censoring times for the lth realization of the kth failure type within the ith cluster. The p-dimensional vector Z_{ikl} denotes a set of covariates. Let T_i be the vector $(T_{ikl}, k = 1, \ldots, K; l = 1, \ldots, L)$ with C_i and Z_i defined similarly. Assume that (T_i, C_i, Z_i), $i = 1, \ldots, n$, are independent and identically distributed with T_i and C_i independent given Z_i. Assume that the $|Z_i|$ are bounded almost surely and the censoring is noninformative conditional on Z_i. Write $X_{ikl} = T_{ikl} \bigwedge C_{ikl}$ and $\triangle_{ikl} = I(T_{ikl} \leq C_{ikl})$, where $a \bigwedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. Under right censoring, the observed data are $\{X_{ikl}, \triangle_{ikl}, Z_{ikl},\, i = 1, \ldots, n;\, k = 1, \ldots, K;\, l = 1, \ldots, L\}$.

Under the frailty model of Vaupel, Manton and Stallard (1979), the intracluster dependency is induced by a frailty variable, ξ, which is common to all members of a cluster. Let ξ_i be a value of ξ associated with the ith cluster. Conditional on ξ_i, the failure times within the cluster are independent. The conditional hazard function of T_{ikl} is given by

$$\lim_{h \downarrow 0} h^{-1} Pr(t \leq T_{ikl} < t + h \mid T_{ikl} \geq t, \xi_i, Z_{ikl}) = \xi_i \alpha_{ikl}(t|Z_{ikl}), \tag{1}$$

where $\{\alpha_{ikl}(t|Z_{ikl})\}$ are termed basic hazard functions. Assume that the random effects ξ_i follow a gamma distribution with mean one and variance θ_0. Murphy (1994, 1995) provided a asymptotic theory for the estimates of the dependence parameter θ_0 and the integrated version of the basic hazard functions. Nielsen et al. (1992) studied (1) when the basic hazard functions follow the proportional hazard model and the frailties ξ_i have a gamma distribution. A maximum likelihood estimate was proposed and implemented using the EM algorithm.

The marginal proportional hazards frailty model postulates the intracluster dependency of (1) while formulating the marginal hazard functions $\lambda_{ikl}(t \mid Z_{ikl})$ for T_{ikl} with the following proportional hazards model:

$$\lambda_{ikl}(t \mid Z_{ikl}) = \lambda_{0k}(t) \exp(\beta_0^T Z_{ikl}), \tag{2}$$

where $\lambda_{0k}(\cdot)$, $k = 1, \ldots, K$, are the unspecified baseline functions and β_0 is the regression parameter. As discussed by Spiekerman and Lin (1998), (2) is very flexible, easy to interpret and mathematically appealing. Under the gamma frailty distribution, the marginal hazard functions and the basic functions are related by $\alpha_{ikl}(t|Z_{ikl}) = \exp\{\theta_0 \int_0^t \lambda_{ikl}(s|Z_{ikl})ds\}\lambda_{ikl}(t|Z_{ikl})$. Glidden (2000) proposed a two-stage estimate for the dependence parameter θ_0. The marginal hazard functions $\lambda_{ikl}(t|Z_{ikl})$ can be estimated by the method of Spiekerman and Lin (1998).

The gamma frailty distribution has been widely used on parametric modeling of intracluster dependency because of its simple interpretation and mathematical tractability (Vaupel, Manton and Stallard (1979), Clayton (1978) and Oakes (1982, 1986)), although many other parametric distributions are possible (Hougaard (1986a, 1986b, 1987), Vaupel (1990) and Aalen (1990)). In some settings there is justification for the assumption that the frailty follows a distribution skewed to the right, such as the gamma or lognormal distribution (Aalen (1988)). Under a gamma frailty structure, some key joint quantities can be efficiently estimated (Glidden and Self (1999)). However, the simplicity and efficiency may be offset by the sensitivity of the estimation to the assumed frailty distribution. Shih and Louis (1995a) showed that different frailty distributions induce quite different dependence structure. Therefore, it is necessary to examine the adequacy of the gamma frailty distribution for intracluster dependence when applying the statistical procedures.

Shih and Louis (1995b) proposed a graphical method for assessing the gamma distribution assumption when the basic functions are parametric and do not depend on covariates. Glidden (1999) developed a test for the gamma frailty model without parameterizing the basic hazard functions when covariates are not involved. In this paper, we suggest a two-stage method for checking the adequacy of the marginal proportional hazards gamma frailty model. In the first stage, the goodness of fit test from Spiekerman and Lin (1996) can be applied for checking the marginal proportional hazards assumption. This paper focuses on the second stage, which is to test for the gamma frailty model. We extend Glidden's (1999) idea for checking the semiparametric gamma frailty model to the marginal proportional hazards frailty model. The two-stage testing procedure provides a model check for using the two-stage method of Glidden (2000). A graphical, as well as a numerical, method for checking the gamma frailty distribution is developed. The proposed test is based on the posterior expectation of the frailties given the observable data over time. Its distribution under the assumed model can be approximated through simulating certain zero-mean Gaussian processes. The Monte Carlo simulations show that the proposed test has reasonable sizes and powers.

The rest of this article is organized as follows. In Section 2, the model checking test process is derived from the posterior expectation of the frailties given the observable data over time and a supremum-type test statistic is proposed. The asymptotic properties of the test process are examined. We show that the test process can be approximated by the sum of independent identically distributed (i.i.d.) processes, which provides a basis for simulating the distributions of the test process and the test statistic, under the gamma frailty model. In Section 3, the finite sample performance of the supremum test is studied under some gamma frailty models and through a few specific alternative models. The graphical diagnoses for these models are also demonstrated. In Section 4, the graphical and numerical techniques for model checking are illustrated through two examples. Finally, the relevant asymptotic results and their proofs are sketched in the Appendix.

2. Test for the Gamma Frailty Distribution

2.1. Test statistic

Based on the posterior expectation of the frailty given the observed data, Glidden (1999) proposed model checking of the gamma frailty model for multivariate failure times without considering covariate variables. We extend his idea to model checking for (1) and (2) where covariate effects are modeled through the marginal proportional hazards model. We focus on checking whether the frailty random effects, ξ_i, $i = 1, \ldots, n$, follow a gamma distribution. For identification purposes, we assume that the gamma distribution has mean one and variance θ, denoted by $Gamma(1, \theta)$. The true value of θ is denoted by θ_0.

Let $Y_{ikl}(t) = I(X_{ikl} \geq t)$ and $N_{ikl}(t) = \triangle_{ikl} I(X_{ikl} \leq t)$. The observed data up to time t is represented by the filtration $\mathcal{F}_t = \sigma\{N_{ikl}(s), Y_{ikl}(s), Z_{ikl} : 0 \leq s \leq t;\ i = 1, \cdots, n;\ k = 1, \cdots, K;\ l = 1, \cdots, L\}$. Let $\xi_{i0}(t) = \mathcal{E}(\xi_i | \mathcal{F}_t)$ be the posterior expectation of ξ_i given $\mathcal{F}_t$. Then the $M_i(t) = N_{i\cdot\cdot}(t) - \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^t \xi_{i0}(s-) Y_{ikl}(s) \exp\{\theta_0 \Lambda_{ikl}(s)\}\, d\Lambda_{ikl}(s)$ are $\mathcal{F}_t$-martingales, where $\Lambda_{ikl}(t) = \int_0^t \lambda_{ikl}(s|Z_{ikl})\, ds$ and $N_{i\cdot\cdot}(t)$ is the summation of $N_{ikl}(t)$ over indices l and k; see Nielsen et al. (1992). Throughout the paper, we denote the summation over a subscript by replacing the subscript with "$\cdot$".

Let $R_i(t, \theta_0) = \sum_{k=1}^{K} \sum_{l=1}^{L} \exp\{\theta_0 \Lambda_{0k}(t \wedge X_{ikl}) e^{\beta_0^T Z_{ikl}}\} - KL + 1$. Under the gamma frailty model, by the arguments of Nielsen et al. (1992), conditional on $\mathcal{F}_t$, ξ_i has the gamma distribution with parameters ${\theta_0}^{-1} + N_{i\cdot\cdot}(t)$ and ${\theta_0}^{-1} R_i(t, \theta_0)$. Thus, $\xi_{i0}(t) = \mathcal{E}(\xi_i | \mathcal{F}_t) = \{1 + \theta_0 N_{i\cdot\cdot}(t)\}/R_i(t, \theta_0)$. The processes $\xi_{10}(t), \ldots, \xi_{n0}(t)$ are independent and identically distributed with mean one. Denote the normalized sum of the ξ_{i0}'s by

$$W_n(t) = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{1 + \theta_0 N_{i\cdot\cdot}(t)}{R_i(t; \theta_0)} - 1 \right\}. \tag{3}$$

It follows from Glidden (1999) that $W_n(t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t H_i(s) dM_i(s)$, where the $H_i(t) = \theta_0 / R_i(t; \theta_0)$ are $\mathcal{F}_t$-predictable processes. The process $W_n(\cdot)$ is the sum of n independent and identically distributed $\mathcal{F}_t$-martingales. Standard martingale theory shows that $W_n(\cdot)$ converges to a zero mean Gaussian process. The process $W_n(\cdot)$ given in (3) provides a natural basis for model checking procedures. Replacing the unknown parameters and functions in $W_n(t)$ by their respective sample estimates, we define the test process

$$\hat{W}_n(t) = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{1 + \hat{\theta} N_{i..}(t)}{\hat{R}_i(t, \hat{\theta})} - 1 \right\}, \tag{4}$$

where $\hat{\theta}$ is the Glidden two stage estimate of frailty parameters for the marginal proportional hazards frailty model (Glidden (2000)), $\hat{R}_i(t, \hat{\theta}) = \sum_{k=1}^{K} \sum_{l=1}^{L} \exp \{\hat{\theta} \hat{\Lambda}_{0k}(t \wedge X_{ikl}) e^{\hat{\beta}^T Z_{ikl}}\} - KL + 1$, $\hat{\beta}$ is the maximum likelihood estimate (mle) based on the quasi-partial likelihood function, and $\hat{\Lambda}_{0k}$ is the Aalen-Breslow type estimate (Spiekerman and Lin (1998)). The estimating functions for these estimates are omitted to save space. Note that $\hat{W}_n(t)$ is the normalized difference between the estimates of the posterior mean of frailties and the mean of frailties under the null hypothesis H_0 that the frailties follow a gamma distribution. We propose the following supremum type test statistic to measure the maximum derivation of "the observed" from "the expected" given by

$$S_n = \sup_{0 \le t \le \tau} |\hat{W}_n(t)|, \tag{5}$$

where $\tau > 0$ is some constant, which can be considered as the end of follow-up time. A large value of S_n indicates that the frailties do not follow a gamma distribution. In subsequent subsections, we study the asymptotic distributions for $\hat{W}_n(t), 0 \le t \le \tau$, and S_n and propose a Monte Carlo procedure to estimate the asymptotic critical values for the test statistic S_n.

2.2. Asymptotic representation for $\hat{W}_n(t)$

Assume the regularity conditions (a)$-$(c) stated in the Appendix. We now introduce the following notations for the convenience of derivation. For $k = 1, \ldots, K$ and $r = 0, 1, 2$, let $S_k^{(r)}(\beta, t) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{L} Y_{ikl}(t) e^{\beta^T Z_{ikl}} Z_{ikl}^{\otimes r}$, $E_k(\beta, t) = S_k^{(1)}(\beta, t)/S_k^{(0)}(\beta, t)$ and $V_k(\beta, t) = S_k^{(2)}(\beta, t)/S_k^{(0)}(\beta, t) - E_k(\beta, t)^{\otimes 2}$, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$. Let $s_k^{(r)}(\beta, t) = \mathcal{E}\{S_k^{(r)}(\beta, t)\}$, where $\mathcal{E}$ denotes expectation. Let $e_k(\beta, t)$ and $v_k(\beta, t)$ be similarly defined as $E_k(\beta, t)$ and $V_k(\beta, t)$ by replacing $S_k^{(r)}(\beta, t)$, $r = 0, 1, 2$, with their respective expectations. Define the norm $\|f(t)\| = \sup_{t \in [0, \tau]} |f(t)|$ for a function $f : [0, \tau] \to R$.

In the following, we show that the test process $\hat{W}_n(t)$ can be approximated by the sum of i.i.d. processes. Let

$$\tilde{f}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{R_i(t, \theta_0)} \Big[N_{i..}(t) - \xi_{i0}(t) \sum_{k=1}^{K} \sum_{l=1}^{L} \exp\{\theta_0 \Lambda_{0k}(t \wedge X_{ikl}) e^{\beta^T Z_{ikl}}\}$$

$$\times\Lambda_{0k}(t\wedge X_{ikl})e^{\beta^T Z_{ikl}}\Big],$$

$$\tilde{g}_k(u,t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}\theta_0 R_i^{-1}(t,\theta_0)\xi_{i0}(t)\exp\{\theta_0\Lambda_{0k}(t\wedge X_{ikl})e^{\beta^T Z_{ikl}}\}e^{\beta^T Z_{ikl}}Y_{ikl}(u),$$

$$\tilde{h}(t) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{L}\Big[\theta_0\xi_{i0}(t)R_i^{-1}(t,\theta_0)\exp\{\theta_0\Lambda_{0k}(t\wedge X_{ikl})e^{\beta^T Z_{ikl}}\}$$

$$\times\Lambda_{0k}(t\wedge X_{ikl})e^{\beta^T Z_{ikl}}Z_{ikl}\Big].$$

Replace θ_0, β_0 and $\Lambda_{0k}(t)$ above, respectively, by $\hat{\theta}$, $\hat{\beta}$ and $\hat{\Lambda}_{0k}(t)$ to obtain $\hat{f}(t)$, $\hat{h}(t)$ and $\hat{g}_k(u,t)$. By Example 2.11.14 of van der Vaart and Wellner (1996), under conditions (a)$-$(c), $\tilde{f}(t)$, $\tilde{h}(t)$ and $\tilde{g}_k(u,t)$ converge in probability uniformly to some cadlag functions $f(t)$, $h(t)$ and $g_k(u,t)$ as $n\to\infty$. By the consistency of $\hat{\theta}$, $\hat{\beta}$ and $\hat{\Lambda}_{0k}(\cdot)$ and conditions (a)$-$(c), it also follows that $\hat{f}(t)$, $\hat{h}(t)$ and $\hat{g}_k(u,t)$ converge in probability uniformly to $f(t)$, $h(t)$ and $g_k(u,t)$ as $n\to\infty$.

By Theorem 2 of Spiekerman and Lin (1998) and some probability arguments similar to Glidden (1999), we have $\hat{W}_n(t) = W_n^0(t) + o_p(1)$, uniformly in $t\in[0,\tau]$, where

$$W_n^0(t) = W_n(t) + n^{1/2}f(t)(\hat{\theta}-\theta_0) + n^{1/2}\sum_{k=1}^{K}\int_0^t g_k(u,t)d\{\hat{\Lambda}_{0k}(u)-\Lambda_{0k}(u)\}$$
$$+n^{1/2}h^T(t)(\hat{\beta}-\beta_0). \tag{6}$$

Now, we note that each of the terms $n^{1/2}(\hat{\theta}-\theta_0)$, $n^{1/2}(\hat{\beta}-\beta_0)$ and $n^{1/2}(\hat{\Lambda}_{0k}(t)$ $-\Lambda_{0k}(t))$ can be approximated by the sums of i.i.d. random processes according to Spiekerman and Lin (1998) and Glidden (2000). Let $M_{ikl}(t) = N_{ikl}(t) - \int_0^t Y_{ikl}(u)\lambda_{0k}(u)e^{\beta_0^T Z_{ikl}(u)}du$ and $w_{i..} = \sum_{k=1}^{K}\sum_{l=1}^{L}\int_0^\tau\{Z_{ikl}-e_k(\beta_0,u)\}dM_{ikl}(u)$. Then $n^{1/2}(\hat{\beta}-\beta_0) = \Sigma_1^{-1}n^{-1/2}\sum_{i=1}^{n}w_{i..}+o_p(1)$, where $\Sigma_1 = \sum_{k=1}^{K}\int_0^\tau v_k(\beta_0,u)s_k^{(0)}$ $(\beta_0,u)\lambda_{0k}(u)du$. It also follows that $n^{1/2}[\hat{\Lambda}_{0k}(t)-\Lambda_{0k}(t)] = n^{-1/2}\sum_{i=1}^{n}\Psi_{ik}(t) + o_p(1)$, where $\Psi_{ik}(t) = \int_0^t(s_k^{(0)}(\beta_0,u))^{-1}dM_{ik.}(u) + r_k(t)^T\Sigma_1{}^{-1}w_{i..}$ and $r_k(t) = -\int_0^t e_k(\beta_0,u)\lambda_{0k}(u)du$.

Let $H_{ikl} = \int_0^\tau Y_{ikl}(u)e^{\beta_0^T Z_{ikl}}d\Lambda_{0k}(u)$ and $R_i(\theta) = \sum_{k=1}^{K}\sum_{l=1}^{L}e^{\theta H_{ikl}} - KL+1$. The $\hat{H}_{ikl}$ and $\hat{R}_i(\theta)$ are obtained by replacing β_0 with $\hat{\beta}$ and $\Lambda_{0k}(t)$ with $\hat{\Lambda}_{0k}(t)$. The pseudo log-likelihood function for θ is given by

$$\hat{l}_n(\theta) = n^{-1}\sum_{i=1}^{n}\Big[\int_0^\tau\log(1+\theta N_{i..}(u))dN_{i..}(u) + \sum_{k=1}^{K}\sum_{l=1}^{L}\theta N_{ikl}(\tau)\hat{H}_{ikl}$$
$$-\{\theta^{-1}+N_{i..}(\tau)\}\log(\hat{R}_i(\theta))\Big]$$

for $\theta \neq 0$, and $\hat{l}_n(0)$ is defined as the limit of $\hat{l}_n(\theta)$ as $\theta \to 0$, which is $-n^{-1}\sum_{i=1}^n \sum_{k=1}^K \hat{H}_{ikl}$. The estimate $\hat{\theta}$ of θ_0 maximizes $\hat{l}_n(\theta)$. Let $I_n(\theta)$ be the negative second derivative of the log-likelihood $\hat{l}_n(\theta)$ with respect to θ, and let $I(\theta)$ be its limit as $n \to \infty$. Then $\sqrt{n}(\hat{\theta} - \theta_0) = n^{-1/2} I^{-1}(\theta_0)\sum_{i=1}^n r_i + O_p(1)$, where $\Upsilon_i = \varepsilon_i(\theta_0) + \sum_{k=1}^K \int_0^\tau \pi_k(u) d\Psi_{ik}(u) + F^T \Sigma_1^{-1} w_{i..}$,

$$\begin{aligned}\varepsilon_i(\theta_0) = &\int_0^\tau \frac{N_{i..}(u-)}{1+\theta_0 N_{i..}(u-)} dN_{i..}(u) + \sum_{k=1}^K \sum_{l=1}^L N_{ikl}(\tau) H_{ikl}\\ &-\{\theta_0^{-1} + N_{i..}(\tau)\} R_i^{-1}(\theta_0) \sum_{k=1}^K \sum_{l=1}^L H_{ikl} e^{\theta_0 H_{ikl}} + \theta^{-2}\log\{R_i(\theta_0)\},\end{aligned}$$

for $\theta_0 \neq 0$, and $\varepsilon_i(0)$ is the limit of $\varepsilon_i(\theta_0)$ as θ_0 goes to 0. The $\pi_k(t)$ and F are defined in Glidden (2000, p.147) as the limits in probability of the $\tilde{\pi}_k(t)$ and $\tilde{F}$ in the following:

$$\begin{aligned}\tilde{\pi}_k(t) = n^{-1} &\sum_{i=1}^n \sum_{l=1}^L e^{\beta^T Z_{ikl}} Y_{ikl}(t) \Big[-\{\theta_0^{-1} + N_{i..}(\tau)\} R_i^{-1}(\theta_0)(1+\theta_0 H_{ikl}) e^{\theta_0 H_{ikl}}\\ &+\theta_0^{-1} R_i^{-1}(\theta_0) e^{\theta_0 H_{ikl}} + N_{ikl}(\tau)\\ &+\{1+\theta_0 N_{i..}(\tau)\} e^{\theta_0 H_{ikl}} R_i^{-2}(\theta_0) \{\sum_{k=1}^K \sum_{l=1}^L H_{ikl} e^{\theta_0 H_{ikl}}\}\Big],\\ \tilde{F} = n^{-1} &\sum_{i=1}^n \sum_{k=1}^K \sum_{l=1}^L Z_{ikl} H_{ikl} \Big[-\{\theta_0^{-1} + N_{i..}(\tau)\} R_i^{-1}(\theta_0)(1+\theta_0 H_{ikl}) e^{\theta_0 H_{ikl}}\\ &+\theta_0^{-1} R_i^{-1}(\theta_0) e^{\theta_0 H_{ikl}} + N_{ikl}(\tau)\\ &+\{1+\theta_0 N_{i..}(\tau)\} e^{\theta_0 H_{ikl}} R_i^{-2}(\theta_0) \{\sum_{k=1}^K \sum_{l=1}^L H_{ikl} e^{\theta_0 H_{ikl}}\}\Big].\end{aligned}$$

Again, the values of $\tilde{\pi}_k(t)$ and $\tilde{F}$ at $\theta_0 = 0$ are defined as their respective limits as $\theta_0 \to 0$. Replacing β, θ_0 and $\Lambda_{0k}(t)$ by $\hat{\beta}$, $\hat{\theta}$ and $\hat{\Lambda}_{0k}(t)$ in $\tilde{\pi}_k(t)$ and $\tilde{F}$ to obtain $\hat{\pi}_k(t)$ and $\hat{F}$, respectively. Some simple probability arguments show that $\hat{\pi}_k(t)$ and $\hat{F}$ are consistent estimates of $\pi_k(t)$ and F.

Note that $W_n(t)$ is the sum of independent and identically distributed $\mathcal{F}_t$-martingales. From previous arguments, we have

$$\hat{W}_n(t) = n^{-1/2} \sum_{i=1}^n \Phi_i(t) + o_p(1), \tag{7}$$

$$\Phi_i(t) = \int_0^t H_i(u) dM_i(u) + f(t) I^{-1}(\theta_0)\Upsilon_i + \sum_{k=1}^K \int_0^t g_k(u,t) d\Psi_{ik}(u) + h^T(t)\Sigma_1^{-1} w_{i..}. \tag{8}$$

It can be shown that the finite dimensional distributions of $\hat{W}_n(t)$ converge to those of a Gaussian process. Along with the tightness of $\hat{W}_n(t)$, the weak convergence of $\hat{W}_n(t)$ to a mean zero Gaussian process $\mathcal{G}(t)$ can be established; see the Appendix for the details.

2.3 Test procedure

Based on the weak convergence of the test process $\hat{W}_n(t)$ to the mean zero Gaussian process $\mathcal{G}(t)$, $0 \le t \le \tau$, it follows from the Continuous Mapping Theorem that $S_n = \sup_{0\le t\le\tau} |\hat{W}_n(t)| \xrightarrow{d} \sup_{0\le t\le\tau} |\mathcal{G}(t)|$ under H_0 as $n \to \infty$. Since the limiting distribution of the test statistic S_n under H_0 is very complicated, we apply a similar simulation idea of Lin, Wei and Ying (1993) to approximate the limiting distribution of $\hat{W}_n(t)$, and thus the critical values of the test.

First, we propose a consistent estimate for the covariance function Cov $\{\mathcal{G}(t), \mathcal{G}(s)\}$ for $s, t \in [0, \tau]$. Notice that the covariance function is equal to $\mathcal{E}\{\Phi_1(t)\Phi_1(s)\}$. Replace the unknown functions and parameters by their corresponding estimates in $\Phi_i(t)$ to get $\hat{\Phi}_i(t)$. Specifically, this involves the following replacements: $f(t)$, $g_k(u,t)$, $h(t)$, $\pi_k(t)$, F, $I(\theta_0)$ by $\hat{f}(t)$, $\hat{g}_k(u,t)$, $\hat{h}(t)$, $\hat{\pi}_k(t)$, $\hat{F}$, $I_n(\hat{\theta})$, respectively; and β, θ_0, $\Lambda_{0k}(t)$, $s_k^{(r)}(t,\beta)$ for $r = 0,1,2$, $e_k(t,\beta)$ and $v_k(t,\beta)$ by $\hat{\beta}$, $\hat{\theta}$, $\hat{\Lambda}_{0k}(t)$, $S_k^{(r)}(t,\hat{\beta})$ for $r = 0,1,2$, $E_k(t,\hat{\beta})$ and $V_k(t,\hat{\beta})$, respectively. We show in the Appendix that $n^{-1}\sum_{i=1}^n \hat{\Phi}_i(t)\hat{\Phi}_i(s)$ is a consistent estimate of the covariance function $\mathcal{E}\{\Phi_1(t)\Phi_1(s)\}$.

Let $\{G_1, \ldots, G_n\}$ be independent standard normal random variables, independent of the original multivariate failure time data. Define

$$\tilde{W}_n(t) = n^{-1/2}\sum_{i=1}^n \hat{\Phi}_i(t)G_i. \tag{9}$$

It follows that $\tilde{W}_n(t)$ given on the observed multivariate failure time data is a Gaussian process with covariance at the times s, t equal to $n^{-1}\sum_{i=1}^n \hat{\Phi}_i(t)\hat{\Phi}_i(s)$, which converges to the asymptotic covariance function of $\hat{W}_n(t)$. Some details are given in the Appendix that, under H_0, the conditional distribution of $\tilde{W}_n(t)$ given the observed data approximates the distribution of $\hat{W}_n(t)$ as $n \to \infty$.

Thus, to approximate the null distribution of $\hat{W}_n(t)$, $0 \le t \le \tau$, we can obtain a number of simulated realizations from $\tilde{W}_n(t)$, $0 \le t \le \tau$, by repeatedly generating independent normal random samples $\{G_i, i = 1, \ldots, n\}$ while holding the observed data fixed. Lack of fit of the hypothesized gamma frailty model could be checked graphically by plotting $\hat{W}_n(t)$ along with a number of realizations, say 20, from $\tilde{W}_n(t)$ conditional on the observed data. Any unusual pattern from the plots of $\hat{W}_n(t)$ in comparing with those from $\tilde{W}_n(t)$, would suggest a lack of fit.

A formal numerical test based on the test statistic S_n given in (5) can be obtained by using the Monte Carlo simulation technique to approximate its critical values. Let $\tilde{S}_n = \sup_{t\in[0,\tau]} |\tilde{W}_n(t)|$, and $c_n(\alpha)$ be the $(1-\alpha)$th sample quantile calculated from B replications of $\tilde{S}_n$, each of which is obtained by generating i.i.d. standard normal random variables. The value of $c_n(\alpha)$ depends on the observed data. The proposed test rejects the null hypothesis H_0 if $S_n > c_n(\alpha)$. Note that $c_n(\alpha)$ converges in probability to the $(1-\alpha)$th quantile of $\sup_{0\le t\le\tau} |\mathcal{G}(t)|$ under H_0 when the sample sizes n and B both tend to infinity. This implies that the significance level of the proposed test converges to its nominal level α. In practice, we recommend $B \ge 1000$ for approximation of a critical value.

3. Numerical Studies

In this section, we carry out a Monte Carlo study to assess the finite sample performance of the proposed test for several models with various sample sizes and different censorship. The empirical sizes and powers are calculated based on 1000 repetitions. The critical values of the test S_n is estimated from 500 simulated realizations from $\tilde{W}_n(t)$.

To examining the size of the proposed test, we consider two null models. For the first model, we assume that the covariate is discrete, and the failure times within each cluster share a common baseline hazard function, i.e., $K = 1$. The number of failure times within each cluster is $L = 2$. Specifically, we let the two failure times within a cluster share the common marginal baseline function $\lambda_{01}(t) = 1$ and let $\beta_0 = 1$. The covariates are given by $Z_{i11} = 0$ and $Z_{i12} = 1$. If the frailty follows the gamma distribution with mean 1 and variance θ_0, then the joint survival function for the failure times (T_{i11}, T_{i12}) is $S(t_1, t_2; \theta_0) = \{\exp(\theta_0 t_1) + \exp(\theta_0 e^1 t_2) - 1\}^{-\theta_0^{-1}}$.

The second model has continuous covariates with $L = 1$ and $K = 2$, i.e., the two failure times in each cluster have different baseline hazard functions. Assume that $\lambda_{01} = 1$, $\lambda_{02} = 2$ and $\beta_0 = 1$, and that the continuous covariates Z_{ikl}, $(i = 1, \ldots, n; k = 1, 2; l = 1)$ are generated from the uniform distribution on $[0, 1]$. Then the marginal hazards for T_{i11} and T_{i21} are $\lambda_{i11} = e^{Z_{i11}}$ and $\lambda_{i21} = 2e^{Z_{i21}}$, respectively. Under the gamma frailty, the joint survivor function of T_{i11} and T_{i21} is $S(t_1, t_2; \theta_0) = \left[\exp\{\theta_0 e^{Z_{i11}} t_1\} + \exp\{2\theta_0 e^{Z_{i21}} t_2\} - 1\right]^{-\theta_0^{-1}}$.

The empirical sizes of the proposed test are given in Table 1 under gamma frailty model for three different sample sizes $(n = 50, 100, 200)$, three different degrees of association $(\theta_0 = 2, 4, 6)$ and two types of censorship (0% and 20%).

To check the power of the test, we consider two alternative models with the same marginal models as the two null models described, except that the frailties are generated from a positive stable distribution with parameter a, say $P(a)$, given by the Laplace transform $E\{\exp(-s\xi)\} = \exp(-s^a)$. The powers of the proposed test are also given in Table 1 under the stable model for three different sample sizes ($n = 50, 100, 200$), three different values of a ($a = 0.1, 0.2, 0.3$) and two types of censorship (0% and 20%).

Table 1. Empirical sizes/power of the proposed test under different models at nominal level $\alpha = 0.05$.

Sample size		$n = 50$		$n = 100$		$n = 200$	
Censoring		0%	20%	0%	20%	0%	20%
Gamma frailty model		Size of test					
Discrete	θ_0=2.0	0.017	0.019	0.020	0.011	0.019	0.020
$K = 1, L = 2$	θ_0=4.0	0.025	0.028	0.041	0.032	0.029	0.020
	θ_0=6.0	0.042	0.034	0.056	0.044	0.042	0.030
Continuous	θ_0=2.0	0.012	0.012	0.016	0.016	0.013	0.026
$K = 2, L = 1$	θ_0=4.0	0.053	0.025	0.051	0.031	0.040	0.025
	θ_0=6.0	0.068	0.034	0.079	0.053	0.090	0.048
Positive stable model		Power of test					
Discrete	a_0=0.1	0.622	0.411	0.919	0.708	0.998	0.922
$K = 1, L = 2$	a_0=0.2	0.590	0.417	0.918	0.709	1.000	0.949
	a_0=0.3	0.479	0.341	0.827	0.588	0.994	0.881
Continuous	a_0=0.1	0.744	0.440	0.996	0.924	1.000	0.999
$K = 2, L = 1$	a_0=0.2	0.808	0.582	0.993	0.946	1.000	1.000
	a_0=0.3	0.643	0.437	0.953	0.831	1.000	0.993

The proposed tests have acceptable sizes in most settings, although they tend to be more conservative when θ_0 is small. The powers are reasonable in both cases when the covariates are chosen to be discrete or continuous. In a simulation study not reported here, we find that as L gets larger, more information is contained in each cluster for the unknown parameters and both empirical size and power of the test improve significantly.

The plots of $\hat{W}_n(t)$ for a typical sample of $n = 100$ from each of the two null models, with $\beta_0 = 1.0$, $\theta_0 = 2$ and 20% of censoring, along with 20 realizations from $\tilde{W}_n(t)$ are given in Figure 1(a) and Figure 1(b), respectively. The p-values shown on these plots are calculated based on 1,000 simulated realizations from $\tilde{W}_n(t)$. These p-values along with the plots indicate that the gamma frailty model is suitable for describing the dependence structure in the sample data. The plots of $\hat{W}_n(t)$ for a typical sample of $n = 100$ from each of the two alternative models, with $\beta_0 = 1.0$, $a = 0.3$ and 20% of censoring, along with 20 realizations from $\tilde{W}_n(t)$, are given in Figures 2(a) and Figure 2(b), respectively. Each of these figures shows clearly that the observed test process departs significantly from the 20 realizations of $\tilde{W}_n(t)$ which, along with the p-value, indicates that the dependence structure within the clusters is not induced by the gamma frailty model.

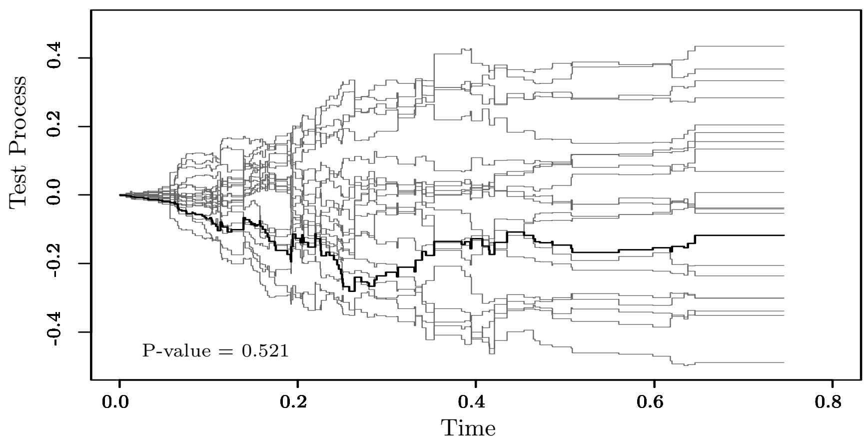

Figure 1(a). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for a typical sample from the first null model with discrete covariate for $K = 1$, $L = 2$, $\theta_0 = 2.0$, $\beta_0 = 1.0$ and sample size $n = 100$.

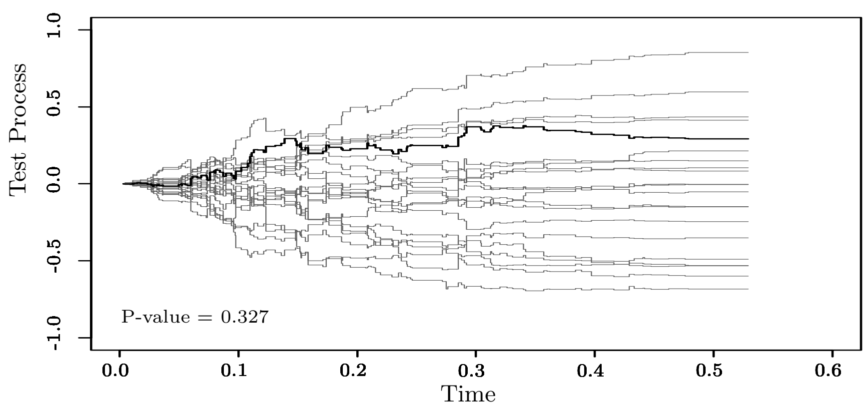

Figure 1(b). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for a typical sample from the second null model with continuous covariate for $K = 2$, $L = 1$, $\theta_0 = 2.0$, $\beta_0 = 1.0$ and sample size $n = 100$.

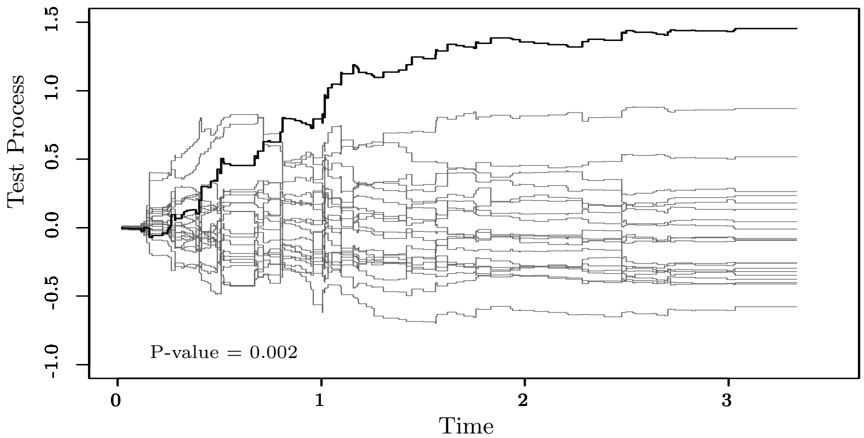

Figure 2(a). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for a typical sample from the first alternative model with discrete covariate for $K = 1$, $L = 2$, $a_0 = 0.3$, $\beta_0 = 1.0$ and sample size $n = 100$.

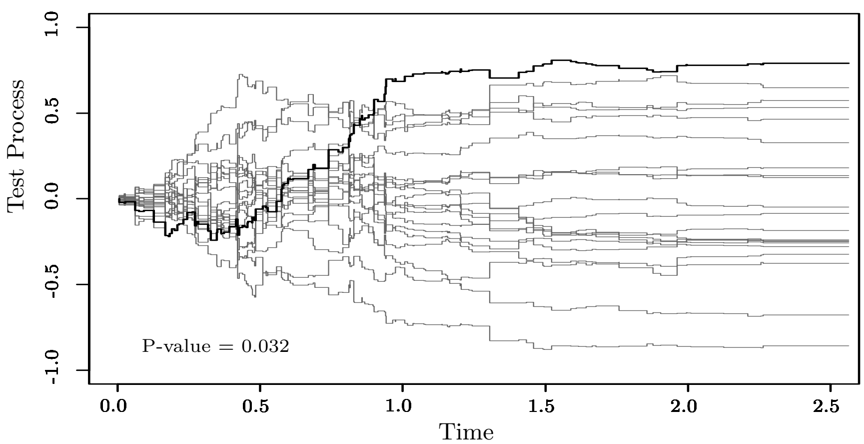

Figure 2(b). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for a typical sample from the second alternative model with continuous covariate for $K = 2$, $L = 1$, $a_0 = 0.3$, $\beta_0 = 1.0$ and sample size $n = 100$.

4. Examples

In this section, we apply our model checking methods to two previously analyzed data sets in the literature. The main interest lies in checking the goodness of fit of the marginal proportional hazards frailty model, and studying a possible correlation between the individuals under study. In our illustration, the p-value for the supremum-type test is always based on 1,000 simulated samples from

$\sup_{t\in[0,\tau]} |\tilde{W}_n(t)|$. In each graphical display, the observed process $\hat{W}_n(t)$ is indicated by a solid curve and 20 simulated realizations from $\tilde{W}_n(t)$ are plotted in dotted grey curves. The p-values for the supremum test are shown on the graph.

4.1. The schizophrenia study

Dr. Ann E. Pulver of Johns Hopkins University conducted a genetic epidemiological study of schizophrenia (Pulver and Liang (1991)). In the study, 487 first degree relatives (273 males, 214 females) of 93 female schizophrenic probands were enrolled. The number of relatives of a single proband ranges from 1 to 12. There are 31 episodes of affective illness (depression or mania or both) out of the 487 relatives in the current database. An important question is whether the risk of affective illness in the relatives is associated with the age at onset of schizophrenia of the proband.

Lin (1994) analyzed this data set using the marginal proportional hazards model where two covariates are considered. One covariate is the proband's age which is dichotomized at 16 years, the other is gender. From the result of parameter estimation, the proband's age at onset is not significant whereas the relative's gender is. Glidden (2000) fit the data using the marginal proportional frailty model. We perform a goodness-of-fit test of the marginal proportional frailty model for the schizophrenia data.

The failure time here is the age at diagnosis of affective illness for the relative. Among relatives of the same proband, the failure times are expected to be correlated. We assume that gender is the only characteristic that differentiates relatives of the same proband. It is natural to define the covariates Z_{ikl} ($i = 1, \ldots, 93$; $k = 1$; $l = 1, \ldots, 12$) as

$$Z_{ikl} = \begin{cases} 1 & \text{if the } l\text{th relative to the } i\text{th proband is male} \\ 0 & \text{if the } l\text{th relative to the } i\text{th proband is female .} \end{cases}$$

We fit the marginal proportional hazards frailty model with a common marginal baseline hazard among all the relatives. The estimate of θ_0 is 1.416 with an estimated standard error 1.091, the estimate of β_0 is -1.239 with an estimated standard error 0.408. The p-value of the supremum test is equal to 0.609. Figure 3(a) displays the observed test process $\hat{W}_n(t)$ along with 20 simulated realizations from $\tilde{W}_n(t)$. Neither graphical nor numerical results provide any evidence against the marginal proportional hazards frailty model.

4.2. The litter-matched tumorigenesis experiment

The study published by Mantel, Bohidar and Ciminera (1977), see also Mantel and Ciminera (1979), is a litter-matched tumorigenesis experiment with one drug treated rat and two placebo treated rats per litter, 50 female litters and 50 male litters. One might expect that the risk of tumor formation depends on the genetic background or the early environmental conditions shared within litters, but differing between litters. There could be an intralitter correlation in time to tumor appearance, the event of interest. Times are given in weeks and deaths before tumor occurrence and yield right-censored observations. In a subset of 50 female litters in this experiment, 40 out of 150 female rats developed a tumor.

Nielsen et al. (1992) used the EM algorithm with the whole data set to estimate the parameters in the proportional hazards frailty model. Restricting attention to the female rates, Hougaard (1986a) analyzed the data based on both a parametric model with Weibull margins and a Cox type model while Clayton (1991) used Gibb's sampling.

We fit the marginal proportional hazards frailty model with $K = 1$, $L = 3$ and a common baseline hazard among all the individuals in each cluster. The covariates Z_{ikl} ($i = 1, \ldots, 50$; $k = 1$; $l = 1, 2, 3$) are given by

$$Z_{ikl} = \begin{cases} 1 & \text{if the rat has the drug treatment} \\ 0 & \text{if the rat has the placebo treatment.} \end{cases}$$

The two-stage estimate of θ_0 is 0.888 with an estimated standard error of 0.519. The estimate of β_0 is 0.856 with an estimated standard error of 0.297. The p-value is 0.308. Figure 3(b) shows the plot of the observed test $\hat{W}_n(t)$ versus 20 simulated realizations from $\tilde{W}_n(t)$. Neither the p-value nor the plot indicate lack-of-fit of the marginal proportional hazards frailty model.

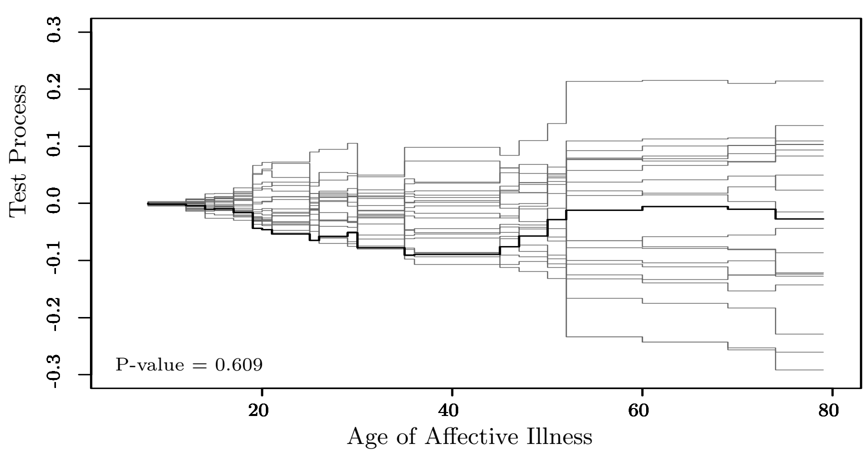

Figure 3(a). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for the schizophrenia study data.

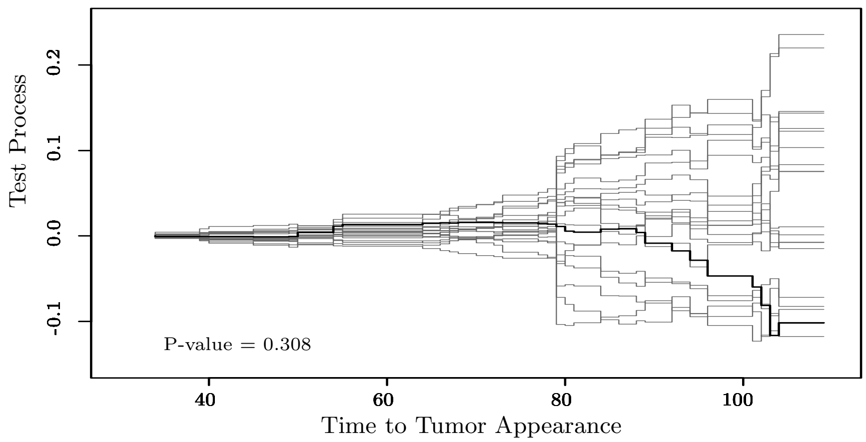

Figure 3(b). Graphical displays of $\hat{W}_n(t)$ versus 20 realizations of $\tilde{W}_n(t)$ for the litter-matched tumorigenesis study data.

Appendix

We present some asymptotic results of the proposed test procedure with proofs. The following additional conditions are assumed. For some $\tau > 0$: (a) $P\{X_{ikl} \geq \tau\} > 0$ for all i, k and l; (b) each component of the variate Z_{ikl} is bounded almost surely for all i, k and l; (c) $\Sigma_1 = \sum_{k=1}^{K} \int_0^\tau v_k(\beta_0, u) s_k^{(0)}(\beta_0, u)\lambda_{0k}(u)\, du$ is positive definite. The conditions (a) to (c) with the i.i.d. assumption imply the conditions a.-f. of Spiekerman and Lin (1998).

Let $D[0,\tau]$ be the set of all uniformly bounded real-valued functions on $[0,\tau]$, endowed with the uniform metric. In the following, we state the relevant asymptotic results and provide a sketch of the proofs.

Theorem 1. *Under the gamma frailty model, $\hat{W}_n(t)$ converges weakly to a mean zero Gaussian process $\mathcal{G}(t)$ in $D[0,\tau]$, as $n \to \infty$, with covariance function* $\mathrm{Cov}\{\mathcal{G}(t), \mathcal{G}(s)\} = \mathcal{E}\{\Phi_1(t)\Phi_1(s)\}$ *for $s, t \in [0,\tau]$, where $\Phi_1(t)$ is given in* (8).

Proof. From the arguments in Section 2, $\hat{W}_n(t)$, $0 \leq t \leq \tau$, is asymptotically equivalent to $W_n^0(t)$, $0 \leq t \leq \tau$, given in (6). So it is sufficient to show the weak convergence of $W_n^0(t)$, $0 \leq t \leq \tau$, which follows from the convergence of the finite-dimensional distributions of $W_n^0(t)$ to those of $\mathcal{G}(t)$ and the tightness of $W_n^0(t)$ by Theorem V.3 (Pollard (1984, p.92)).

By (7), (8) and the application of the Multivariate Central Limit Theorem (Billingsley (1995, p.357)), the finite-dimensional distributions of the process $W_n^0(t)$, $0 \leq t \leq \tau$, converge to those of $\mathcal{G}(t)$, $0 \leq t \leq \tau$.

By the definition of tightness under the uniform metric, the tightness of $W_n^0(t)$, $0 \leq t \leq \tau$, will follow from the tightness of each of the following terms of

(6) on $0 \le t \le \tau$. Let $Q_1(t) = n^{-1/2}\sum_{i=1}^n \int_0^t H_i(s)dM_i(s)$, $Q_2(t) = n^{1/2}f(t)(\hat{\theta} - \theta_0)$, $Q_{3k}(t) = n^{1/2}\int_0^t g_k(s,t)d\{\hat{\Lambda}_{0k}(s) - \Lambda_{0k}(s)\}$ and $Q_4(t) = n^{1/2}h^T(t)(\hat{\beta} - \beta_0)$. By Theorem VIII.13 (Pollard (1984, p.179)), we have the weak convergence of $Q_1(t)$, which implies the tightness of $Q_1(t)$.

It can be easily checked that the functions $f(t)$, $h(t)$ and $g_k(s,t)$ are cadlag functions on $0 \le s \le \tau$, $0 \le t \le \tau$. The tightness of $Q_2(t)$ and of $Q_4(t)$ follows from the weak convergence of $n^{1/2}(\hat{\theta} - \theta_0)$ and $n^{1/2}(\hat{\beta} - \beta_0)$. Simple calculation also shows that the functions $g_k(s,t), (s,t) \in [0,\tau]^2$ for $1 \le k \le K$ have bounded variations. By the weak convergence of $n^{1/2}(\hat{\Lambda}_{0k}(t) - \Lambda_{0k}(t))$, $0 \le t \le \tau$, and the Skorohod-Dudley-Wichura representation (Shorack and Wellner (1986, p.47)), we have the weak convergence of $Q_{3k}(t)$, therefore the tightness of $Q_{3k}(t)$, on $0 \le t \le \tau$.

Theorem 2. *Under the gamma frailty model, $n^{-1}\sum_{i=1}^n \hat{\Phi}_i(t)\hat{\Phi}_i(s)$ converges in probability to $\mathcal{E}\{\Phi_1(t)\Phi_1(s)\}$ as $n \to \infty$, uniformly for $s,t \in [0,\tau]$.*

Proof. By Example 2.11.14 of van der Vaart and Wellner (1996), under the conditions (a)$-$(c), $n^{-1}\sum_{i=1}^n \Phi_i(t)\Phi_i(s) \xrightarrow{P} \mathcal{E}\{\Phi_1(t)\Phi_1(s)\}$, uniformly for $s,t \in [0,\tau]$. So it suffices to show that $|n^{-1}\sum_{i=1}^n \hat{\Phi}_i(t)\hat{\Phi}_i(s) - n^{-1}\sum_{i=1}^n \Phi_i(t)\Phi_i(s)| \to 0$ in probability as $n \to \infty$, uniformly for $s,t \in [0,\tau]$. Note that

$$\sum_{i=1}^n\{\hat{\Phi}_i(t)\hat{\Phi}_i(s) - \Phi_i(t)\Phi_i(s)\} = \sum_{i=1}^n\{\hat{\Phi}_i(t) - \Phi_i(t)\}\hat{\Phi}_i(s) + \sum_{i=1}^n\{\hat{\Phi}_i(s) - \Phi_i(s)\}\Phi_i(t).$$

It remains to show that

$$\begin{aligned} &\max_{1\le i\le n}\sup_{0\le t\le\tau} |\hat{\Phi}_i(t) - \Phi_i(t)| = o_p(1), \\ &\max_{1\le i\le n}\sup_{0\le t\le\tau} |\hat{\Phi}_i(t)| = O_p(1), \qquad (10) \\ &\max_{1\le i\le n}\sup_{0\le t\le\tau} |\Phi_i(t)| = O_p(1). \end{aligned}$$

Examining each term of $\Phi_i(t)$ given in (8) and each term of $\hat{\Phi}_i(t)$ as described following (8), by the consistency of $\hat{\theta}$, $\hat{\beta}$, $\hat{\Lambda}_{0k}$, and $I_n(\hat{\theta})$ and the conditions (a)$-$(c), (10) follows from some routine probability arguments. We omit the lengthy and tedious details.

Theorem 3. *Under the gamma frailty model, conditional on the observed data sequence, $\tilde{W}_n(t)$ converges weakly to $\mathcal{G}(t)$ in $D[0,\tau]$ as $n \to \infty$.*

Proof. By Theorem 3 and the Multivariate Central Limit Theorem (Billingsley (1995, p.357)), the finite-dimensional distributions of $\tilde{W}_n(t)$, $0 \le t \le \tau$, given the observed data sequence, converge to those of $\mathcal{G}(t)$, $0 \le t \le \tau$. Now we show that the process $\tilde{W}_n(t)$, $0 \leq t \leq \tau$, given the observed data sequence, is also tight. We check the extension to the moment condition of Billingsley (1968, p.128) established by McKeague and Zhang (1994, p.506). For $t_1 \leq t \leq t_2$,

$$\begin{aligned}
&\mathcal{E}\left[\{\tilde{W}_n(t)-\tilde{W}_n(t_1)\}^2\{\tilde{W}_n(t_2)-\tilde{W}_n(t)\}^2 \,\Big|\, \text{data}\right] \\
=\;& 3n^{-2}\sum_{i=1}^{n}\{\hat{\Phi}_i(t)-\hat{\Phi}_i(t_1)\}^2\{\hat{\Phi}_i(t_2)-\hat{\Phi}_i(t)\}^2 \\
&+3n^{-2}\sum_{i,k=1,i\neq k}^{n}\{\hat{\Phi}_i(t)-\hat{\Phi}_i(t_1)\}^2\{\hat{\Phi}_i(t_2)-\hat{\Phi}_i(t)\}^2 \\
\leq\;& 3n^{-1}\sum_{i=1}^{n}\{\hat{\Phi}_i(t)-\hat{\Phi}_i(t_1)\}^2 \;\; n^{-1}\sum_{i=1}^{n}\{\hat{\Phi}_i(t_2)-\hat{\Phi}_i(t)\}^2+o_p(1) \\
\xrightarrow{P}\;& \mathcal{E}\{\Phi_i(t)-\Phi_i(t_1)\}^2 \;\; \mathcal{E}\{\Phi_i(t_2)-\Phi_i(t)\}^2,
\end{aligned}$$

uniformly in $0 \leq t_1 \leq t \leq t_2 \leq \tau$ by (10). By (8),

$$\begin{aligned}
\mathcal{E}\{\Phi_i(t)-\Phi_i(t_1)\}^2 \leq\;& 4\mathcal{E}\{\int_{t_1}^{t} H_i(u)dM_i(u)\}^2+4\{f(t)-f(t_1)\}^2 I^{-2}(\theta_0)\mathcal{E}\Upsilon_i^2 \\
&+4\mathcal{E}\Big[\sum_{k=1}^{K}\{\int_0^t g_k(u,t)d\Psi_{ik}(u)-\int_0^{t_1} g_k(u,t_1)d\Psi_{ik}(u)\}\Big]^2 \\
&+\mathcal{E}[h^T(t)\Sigma_1^{-1}w_{i..}-h^T(t_1)\Sigma_1^{-1}w_{i..}]^2.
\end{aligned}$$

Examining each of the above terms carefully, under the conditions (a)−(c) we can find a nondecreasing, continuous function $\psi(t)$ on $[0,\tau]$ such that $\mathcal{E}\{\Phi_i(t)-\Phi_i(t_1)\}^2 \leq \psi(t)-\psi(t_1)$. This completes the proof.

Acknowledgements

The authors thank David V. Glidden for providing his computer programs and giving helpful suggestions. Yanqing Sun's work was partially supported by the National Science Foundation grant DMS-0304922 and by funds provided by the University of North Carolina at Charlotte. The authors also thank the referees for helpful comments that improved the presentation of the paper.

References

Aalen, O. O. (1988). Heterogeneity in survival analysis. *Statist. Medicine* **7**, 1121-1137.

Aalen, O. O. (1990). Modeling heterogeneity in survival analysis by the compound Poisson distribution. Technical Report, Section of Medical Statistics, University of Oslo.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Billingsley, P. (1995). *Probability and Measure*. Wiley, New York.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with Discussion). *J. Roy. Statist. Soc. Ser. A* **148**, 82-117.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467-485.

Glidden, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika* **86**, 381-393.

Glidden, D. V. and Self, S. G. (1999). Semiparametric likelihood estimation in the Clayton-Oakes failure time model. *Scand. J. Statist.* **26**, 363-372.

Glidden, D. V. (2000). A two-stage estimator of the dependence parameter for the Clayton-Oakes Model. *Lifetime Data Anal.* **6**, 141-156.

Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika* **73**, 671-678.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.

Hougaard, P. (1987). Modeling multivariate survival. *Scand. J. Statist.* **14**, 291-304.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Medicine* **13**, 2233-2247.

Mantel, N., Bohidar, N. R. and Ciminera, J. L. (1977). Mantel-Haenszel analysis of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Research* **37**, 3863-3868.

Mantel, N. and Ciminera, J. L. (1979). Use of logrank scores in the analysis of litter matched data on time to tumor appearance. *Cancer Research* **39**, 4308-4315.

McKeague, I. W. and Zhang, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22**, 495-514

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25-43.

Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* **44**, 414-422.

Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika* **73**, 353-361.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84**, 487-493.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Pulver, A. E. and Liang, K. Y. (1991). Estimating effects of proband characteristics and familial risk: II. the association between age at onset and familial risk in the Maryland schizophrenia sample. *Genetic Epidemiology* **8**, 339-350.

Shih, J. H. and Louis, T. A. (1995a). Inference on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384-1399.

Shih, J. H. and Louis, T. A. (1995b). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Anal.* **1**, 205-220.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Spiekerman, C. F. and Lin, D. Y. (1996). Checking the marginal Cox model for correlated failure time data. *Biometrika* **83**, 143-156.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.* **93**, 1164-1175.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, 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**, 439-454.

Vaupel, J. W. (1990). Kindred lifetimes: frailty models in population genetics. In *Convergent Issues in Genetics and Demography* (Edited by J. Adams, A. Hermalin, D. Lam and P. Smouse). Oxford University Press, Oxford.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Transamerica Occidental Life Insurance Charlotte, NC 28202, U.S.A.

E-mail: sfcui97@msn.com

Department of Mathematics, The University of North Carolina at Charlotte, Charlotte, NC 28223, U.S.A.

E-mail: yasun@uncc.edu

(Received July 2002; accepted February 2003)