

Pairwise dependence diagnostics for clustered failure-time data

BY DAVID V. GLIDDEN

*Department of Epidemiology and Biostatistics, University of California, 185 Berry Street,
 Lobby 4, Suite 5700, San Francisco, California 94107-1762, U.S.A.*

dave@biostat.ucsf.edu

SUMMARY

Frailty and copula models specify a parametric dependence structure for multivariate failure-time data. Estimation of some joint quantities can be highly sensitive to the assumed parametric form, and hence model fit is an important issue. This paper lays out a general diagnostic framework for evaluating and selecting frailty and copula models. The approach is based on the cumulative sum of residuals that are calculated in bivariate time. The residuals reflect the difference between the observed and expected bivariate association structures. The proposed model-checking process is interpretable with a limiting distribution which can be approximated using the bootstrap. Simulations and a data example illustrate the practical application of the method.

Some key words: Bivariate failure-time data; Censoring; Copula model; Cross-ratio function; Frailty model.

1. INTRODUCTION

Multivariate failure-time data are clustered samples of possibly censored time-to-event observations. The clusters may be individuals, families or geographic units. These data occur, for example, in family disease studies where the cluster is a pedigree and the response is the age at onset of a disease.

Analyzing the strength and implications of intra-family disease clustering requires a joint model for the disease onset times. Models can be constructed using two closely related strategies, namely frailty and copula models. Shared frailty models induce dependence using latent random effects, termed frailties, which act multiplicatively on the hazard. The modelling is completed by specifying a parametric family of probability densities, $g_\gamma(\cdot)$, for the frailties. If $p_\gamma(\cdot)$ is the Laplace transformation based the density and $q_\gamma(\cdot)$ equals $p_\gamma^{-1}(\cdot)$, the survival function for failure times (T_1, T_2) is

$$\text{pr}(T_1 > t_1, T_2 > t_2) = p_\gamma \left[\sum_{k=1}^2 q_\gamma \{ \mathcal{S}_k(t_k) \} \right], \quad (1)$$

where $\mathcal{S}_k(t) = \text{pr}(T_k > t)$. Copula models use a parametric family of paired distributions with uniform margins, $\mathcal{C}_\gamma(u_1, u_2)$, to specify distributions with arbitrary margins such that

$$\text{pr}(T_1 > t_1, T_2 > t_2) = \mathcal{C}_\gamma \{ \mathcal{S}_1(t_1), \mathcal{S}_2(t_2) \}. \quad (2)$$

Both copula and frailty models specify a parametric bivariate dependence structure. The form is implied by choice of $g_\gamma(\cdot)$ for a frailty model and by specification of $\mathcal{C}_\gamma(\cdot, \cdot)$ for a copula model. The two approaches are intimately connected (Oakes, 1989). In fact, (1)

defines a copula function $C_\gamma(\cdot, \cdot)$ for a frailty model with a density $g_\gamma(\cdot)$. In this paper, we focus on checking frailty models but application to copula models is straightforward.

Frailty and copula models facilitate the estimation of complex joint quantities. For example, given disease onset times (T_1, T_2) on a set of twins, the conditional risk $G(t; s)$ of failure for one twin, T_2 , given failure of T_1 at time s , may be of interest; here,

$$G(t; s) = \text{pr}(T_2 \leq t | T_1 = s, T_2 > s), t > s.$$

Nonparametric estimation of $G(t; s)$ may be hampered by few failures of T_1 at or near s .

Under a frailty or copula model, $G(t; s)$ can be readily estimated based on the estimates of $S_k(\cdot)$, ($k = 1, 2$) and γ (Nielsen et al., 1992; Shih & Louis, 1995; Hsu & Prentice, 1996a) in expressions (1) or (2). In practice, such estimates can be quite sensitive to the assumed model for dependence. Figure 1 graphs estimates of $G(t; s)$ based on a dataset of 1718 monozygotic twins (Duffy et al., 1990), where (T_1, T_2) represent ages at appendectomy and $s = 3$ years of age. The figure shows that gamma (Clayton, 1978) and positive stable (Hougaard, 1986) frailty models give markedly different estimates of the future risk of an unaffected co-twin. Accurate conclusions will require that an appropriately fitting model be selected. Here it is unclear which model, if either, reflects the dependence structure of the data.

A number of authors have considered tests and graphical procedures for checking the dependence structure of clustered failure-time data. Most of these methods have focused on checking the fit of a particular distribution (Shih & Louis, 1995; Shih, 1998; Glidden, 1999). An interesting class of approaches has been used by Chen & Bandeen-Roche (2005) and by Viswanathan & Manatunga (2001). They compared an empirical to a model-based estimate of Kendall's τ . Asymptotic theory for these approaches is not fully developed.

This paper proposes a general diagnostic approach to check the bivariate association structure of clustered failure times. Checking uses residuals which are based on a measure of bivariate dependence, the cross-ratio function. The residuals compare the assumed to

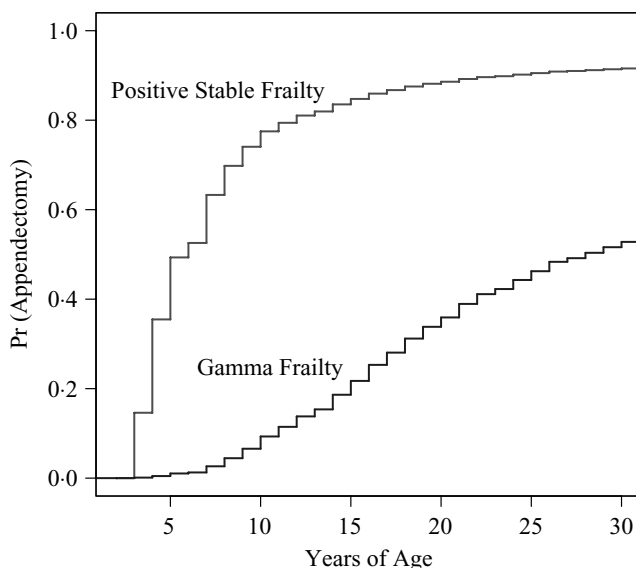


Fig. 1. Australian monozygotic twin data. Conditional risk of appendectomy for co-twin given that the index twin has appendectomy at age 3: results for gamma and positive stable frailty fits.

the empirical cross-ratio function. By inspecting the pattern of the residuals, it is possible to detect a lack of fit and also to deduce the pattern of the departure from the assumed model.

2. MODEL-CHECKING PROCESS

2.1. Data and notation

The method accommodates arbitrary cluster sizes; however, we introduce it for paired failure times and consider issues of covariates and arbitrary cluster sizes in §3.

For $(i = 1, \dots, n)$ let (T_{i1}, T_{i2}) be the i th pair. Failures are right censored by (C_{i1}, C_{i2}) , which is independent of the failure times. The distribution of the failures is continuous with vectors $\{(T_{ik}, C_{ik}), k = 1, 2\}$ ($i = 1, \dots, n$) independent and identically distributed. The observed data are $X_{ik} = \min(T_{ik}, C_{ik})$ and $\delta_{ik} = I(T_{ik} \leq C_{ik})$, ($k = 1, 2$) where $I(\cdot)$ is the indicator function. Let $Y_{ik}(t_k) = I(X_{ik} \geq t_k)$ and $N_{ik}(t_k) = I(X_{ik} \leq t_k, \delta_{ik} = 1)$ be the at-risk and counting processes. The k th failure is observed over $[0, \tau_k]$ for all clusters and $\text{pr}(X_1 > \tau_1, \dots, X_K > \tau_K) > 0$. Finally, denote the j th ordered observed failure time of T_k by $t_{k(j)}$ ($j = 1, \dots, D_k$), ($k = 1, 2$).

2.2. The cross-ratio function

The cross-ratio function at time (t_1, t_2) , denoted by $\theta(t_1, t_2)$, was defined by Clayton (1978) as

$$\theta(t_1, t_2) := \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 \geq t_2)} := \frac{\lambda_2(t_2|T_1 = t_1)}{\lambda_2(t_2|T_1 \geq t_1)} := \frac{S(dt_1, dt_2)S(t_1^-, t_2^-)}{S(t_1^-, dt_2)S(dt_1, t_2^-)}, \quad (3)$$

where

$$\lambda_k(t_k|\Omega) = \lim_{h \downarrow 0} h^{-1} \text{pr}(t_k \leq T_k < t_k + h|\Omega, T_k \geq t_k),$$

for the arbitrary event Ω , and $S(t_1, t_2) := \text{pr}(T_1 > t_1, T_2 > t_2)$. The cross-ratio function at (t_1, t_2) is the ratio of the hazard of T_1 at t_1 given $T_2 = t_2$ compared with the hazard of T_1 at t_1 given $T_2 \geq t_2$. The function is symmetric, so that it is also the ratio of the hazard of T_2 at t_2 given $T_1 = t_1$ compared with the hazard given $T_1 \geq t_1$.

The choice of distribution for the frailty imparts a distinctive pattern on the cross-ratio function. For instance, a gamma frailty model specifies a cross-ratio function which is constant in time whereas cross-ratio functions for positive stable and inverse Gaussian frailties decrease with time.

2.3. Three hazard functions

Gill et al. (1995) showed that the probability law of (T_1, T_2) specifies and is specified by three bivariate hazard functions. The functions have the form

$$\begin{aligned} \Lambda_{11}(t_1, t_2) &= \int_0^{t_1} \int_0^{t_2} \frac{S(ds_1, ds_2)}{S(s_1^-, s_2^-)}, \quad \Lambda_{01}(t_1^-, t_2) = \int_0^{t_2} \frac{S(t_1^-, ds_2)}{S(t_1^-, s_2^-)}, \\ \Lambda_{10}(t_1, t_2^-) &= \int_0^{t_1} \frac{S(ds_1, t_2^-)}{S(s_1^-, t_2^-)}. \end{aligned} \quad (4)$$

These are the hazard functions for double and single failures at (t_1, t_2) . By (3), we have

$$\theta(t_1, t_2) = \frac{\Lambda_{11}(dt_1, dt_2)}{\Lambda_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2)},$$

$$\Lambda_{11}(dt_1, dt_2) - \theta(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) = 0. \quad (5)$$

Formula (5) leads to a residual which compares the parametric and nonparametric information in the data about pairwise dependence. The hazard functions can be consistently estimated (Gill et al., 1995) by the empirical Nelson-Aalen-type estimators

$$\begin{aligned} \hat{\Lambda}_{11}(t_1, t_2) &= \int_0^{t_1} \int_0^{t_2} \frac{\sum_{i=1}^n N_{i1}(ds_1)N_{i2}(ds_2)}{\sum_{i=1}^n Y_{i1}(s_1)Y_{i2}(s_2)}, \\ \hat{\Lambda}_{01}(t_1^-, t_2) &= \int_0^{t_2} \frac{\sum_{i=1}^n N_{i2}(ds_2)Y_{i1}(t_1)}{\sum_{i=1}^n Y_{i2}(s_2)Y_{i1}(t_1)}, \\ \hat{\Lambda}_{10}(t_1, t_2^-) &= \int_0^{t_1} \frac{\sum_{i=1}^n N_{i1}(ds_1)Y_{i2}(t_2)}{\sum_{i=1}^n Y_{i1}(s_1)Y_{i2}(t_2)}. \end{aligned} \quad (6)$$

Frailty model diagnostics proceed by considering a candidate frailty distribution for the data. Estimation of the parameters of the frailty model leads to a model-based estimator of the cross-ratio function, denoted by $\hat{\theta}(\cdot)$. Given that estimator, equation (5) defines a series of local residuals. The residual at (t_1, t_2) ,

$$\hat{\Lambda}_{11}(dt_1, dt_2) - \hat{\theta}(t_1, t_2)\hat{\Lambda}_{10}(dt_1, t_2^-)\hat{\Lambda}_{01}(t_1^-, dt_2), \quad (7)$$

takes values on the grid of ordered failure times $\{t_{1(j)}, t_{2(l)}\}$ ($j = 1, \dots, D_1$), ($l = 1, \dots, D_2$). Define the two-by-two table at $\{t_{1(j)}, t_{2(l)}\}$ as

$$\begin{array}{cc} T_1 = t_{1(j)} & T_1 \geq t_{1(j)} \\ C_1 \geq t_{1(j)} & C_1 \geq t_{1(j)} \\ T_2 = t_{2(l)} & \\ C_2 \geq t_{2(l)} & \Psi_{jl} \quad d_{2jl} \\ T_2 \geq t_{2(l)} & \\ C_2 \geq t_{2(l)} & d_{1jl} \quad R_{jl} \end{array}$$

In the table,

$$\begin{aligned} \Psi_{jl} &= \sum_{i=1}^n N_{i1}\{\Delta t_{1(j)}\}N_{i2}\{\Delta t_{2(l)}\}, \quad R_{jl} = \sum_{i=1}^n Y_{i1}\{t_{1(j)}\}Y_{i2}\{t_{2(l)}\}, \\ d_{2jl} &= \sum_{i=1}^n Y_{i1}\{t_{1(j)}\}N_{i2}\{\Delta t_{2(l)}\}, \quad d_{1jl} = \sum_{i=1}^n Y_{i2}\{t_{2(l)}\}N_{i1}\{\Delta t_{1(j)}\}. \end{aligned}$$

Here $N(\Delta t)$ denotes $N(t) - N(t^-)$. The residual (7) is equal to

$$R_{jl}^{-1} \left[\Psi_{jl} - \hat{\theta}\{t_{1(j)}, t_{2(l)}\} \frac{d_{1jl}d_{2jl}}{R_{jl}} \right], \quad (8)$$

which is proportional to the difference between the observed and expected values of Ψ_{jl} when the cross-ratio function equals $\hat{\theta}(\cdot)$.

Statistics based on Ψ have been used in a number of previous methods. For instance, Hsu & Prentice (1996b) formed a test of independence for paired failures using sums of

$\Psi - 1$. The variable Ψ was also the basis for dependence parameter estimation proposed by Clayton (1978) for the gamma frailty model.

We define a model-checking process by considering the cumulative sum of these residuals,

$$\hat{F}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \{\hat{\Lambda}_{11}(ds_1, ds_2) - \hat{\theta}(s_1, s_2)\hat{\Lambda}_{10}(ds_1, s_2^-)\hat{\Lambda}_{01}(s_1^-, ds_2)\}.$$

A weighted version of the process would have the form

$$\hat{F}_W(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} W_n(s_1, s_2)\{\hat{\Lambda}_{11}(ds_1, ds_2) - \hat{\theta}(s_1, s_2)\hat{\Lambda}_{10}(ds_1, s_2^-)\hat{\Lambda}_{01}(s_1^-, ds_2)\}, \quad (9)$$

where $W_n(\cdot)$ is a user-specified weight function of bounded variation which converges uniformly to a bounded, continuous limiting function $W(\cdot)$. Reasonable choices for the weights could include $n^{-1} \sum Y_{i1}(s_1)Y_{i2}(s_2)$ or an estimator of the bivariate survivor function (Dabrowska, 1988; Prentice & Cai, 1992).

The residuals compare the observed to the model-based association. Thus, the pattern of residuals reveals the nature of the departure, in terms of the cross-ratio function, and can suggest better-fitting models.

The model-checking process \hat{F}_W can be used to develop test statistics for formal assessments of model fit. A single point in time (t_1, t_2) can be selected and the test statistic E_t defined, where

$$E_t := \hat{F}_W(t_1, t_2)/\text{ESE}\{\hat{F}_W(t_1, t_2)\},$$

in which ESE denotes the estimated standard error. A supremum-type test statistic could be used, such as

$$Q := \sup_{t_1, t_2} |\hat{F}_W(t_1, t_2)|.$$

3. COVARIATES AND ARBITRARY CLUSTER SIZES

3.1. Multivariate failure time data

Consider observation of a failure vector (T_1, \dots, T_K) , where $K > 2$. We assume that the marginal distributions are identical and also that the data have an exchangeable bivariate association structure. The former assumption will be relaxed in §3.2 and the latter assumption is satisfied for shared frailty and copula models. If the association structure is not exchangeable, one might still use a synthesized model-checking process if association between pairs followed a similar functional form, even if the magnitude varied by pair. Otherwise, one could inspect the series of bivariate associations across the failure vector. Below we give a synthesized model-checking approach assuming a common association structure across the failure vector.

Given the above, pairs from the failure vector have a common bivariate survivor function, such that

$$\text{pr}(T_k > t_1, T_l > t_2) = S(t_1, t_2),$$

for all $k \neq l$. The three hazard functions can be defined as in (4) using the common joint survivor function; their notation includes the superscript K to emphasize the arbitrary size of the failure vector. These hazard functions are symmetric:

$$\Lambda_{11}^K(t_1, t_2) = \Lambda_{11}^K(t_2, t_1), \quad \Lambda_{10}^K(t_1, t_2) = \Lambda_{01}^K(t_2, t_1).$$

Possible estimators include

$$\begin{aligned}\hat{\Lambda}_{11}^K(t_1, t_2) &= \int_0^{t_2} \int_0^{t_1} \frac{\sum_{i=1}^n \sum_{k=1}^K \sum_{l \neq k} N_{ik}(ds_1) N_{il}(ds_2)}{\sum_{i=1}^n \sum_{k \neq l} Y_{ik}(s_1) Y_{il}(s_2)} \\ \hat{\Lambda}_{10}^K(t_1, t_2^-) &= \int_0^{t_1} \frac{\sum_{i=1}^n \sum_{k=1}^K \sum_{l \neq k} N_{ik}(ds_1) Y_{il}(t_2)}{\sum_{i=1}^n \sum_{k \neq l} Y_{ik}(s_1) Y_{il}(t_2)} \\ \hat{\Lambda}_{01}^K(t_1^-, t_2) &= \int_0^{t_2} \frac{\sum_{i=1}^n \sum_{k=1}^K \sum_{l \neq k} N_{ik}(ds_2) Y_{il}(t_1)}{\sum_{i=1}^n \sum_{k \neq l} Y_{ik}(s_2) Y_{il}(t_1)}.\end{aligned}$$

Note that $\hat{\Lambda}_{11}^K(t_1, t_2) = \hat{\Lambda}_{11}^K(t_2, t_1)$ and $\hat{\Lambda}_{10}^K(t_1, t_2^-) = \hat{\Lambda}_{01}^K(t_2^-, t_1)$ because the estimators above permute from among the pair members. Substituting estimators into (9) gives the model-checking function

$$\hat{F}_{1W}^K(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} W_n(s_1, s_2) \{ \hat{\Lambda}_{11}^K(ds_1, ds_2) - \hat{\theta}(s_1, s_2) \hat{\Lambda}_{10}^K(ds_1, s_2^-) \hat{\Lambda}_{01}^K(s_1^-, ds_2) \},$$

for weight function $W_n(\cdot, \cdot)$.

A more ad hoc method might consider the pair (T_k, T_l) chosen from among the $K(K - 1)$ possible pairs from the vector of failures. Based on a given pair, one can define the bivariate hazard functions as $(\hat{\Lambda}_{11}^{kl}, \hat{\Lambda}_{10}^{kl}, \hat{\Lambda}_{01}^{kl})$ and then a model-checking process $\hat{F}_W^{kl}(t_1, t_2)$ which equals

$$\int_0^{t_1} \int_0^{t_2} W^{kl}(s_1, s_2) \{ \hat{\Lambda}_{11}^{kl}(ds_1, ds_2) - \hat{\theta}(s_1, s_2) \hat{\Lambda}_{10}^{kl}(ds_1, s_2^-) \hat{\Lambda}_{01}^{kl}(s_1^-, ds_2) \}.$$

These can be combined to define the process

$$\hat{F}_{2W}^K(t_1, t_2) = \sum_{k=1}^K \sum_{l \neq k} \hat{F}_W^{kl}(t_1, t_2). \quad (10)$$

If the weight functions are symmetric such that $W^{kl}(s_1, s_2)$ equals $W^{lk}(s_2, s_1)$, then (10) is also symmetric.

3.2. Covariates

The marginal distributions may have some type of structure. For instance, a subject's hazard may depend on a series of possibly time-dependent covariates, $Z_{ik}(\cdot)$ ($i = 1, \dots, n$), ($k = 1, \dots, K$). Suppose we relax the censoring assumption to allow T_{ik} to be independent of C_{ik} conditional on the covariates and assume that the vectors $\{[X_{ik}, \delta_{ik}, Z_{ik}(\cdot)], k = 1, \dots, K\}$ ($i = 1, \dots, n$) are independent and identically distributed. Let the marginal hazard function for the k th failure in the i th cluster $\alpha_{ik}(t)$ equal

$$\alpha_{0k}(t) \exp\{\beta Z_{ik}(t)\}.$$

Other specifications are possible and could include $\alpha_{0k}(\cdot)$, $\alpha_0(\cdot) \exp\{\beta Z_{ik}(t)\}$ or fully parametric models. Let $A_{ik}(t)$ be $\int_0^t \alpha_{ik}(s) ds$ with $\mathcal{S}_{ik}(t) := \exp\{-A_{ik}(t)\}$. It is possible to write the data as a copula function $\mathcal{C}_\gamma(\cdot, \cdot)$ such that the joint survivor function for the i th pair is

$$\text{pr}(T_{i1} > t_1, T_{i2} > t_2) = \mathcal{C}_\gamma\{\mathcal{S}_{i1}(t_1), \mathcal{S}_{i2}(t_2)\}$$

($i = 1, \dots, n$).

In the covariate case, pairs share a common association structure on a time scale transformed by the marginal hazard or survival function. Provided that the marginal model is correctly specified, the vectors $[\exp\{-A_{ik}(X_{ik})\}, \delta_{ik}, (k = 1, \dots, K)], (i = 1, \dots, n)$ are independent and identically distributed with uniform marginals, conditional on the covariates, and common cross-ratio function,

$$\theta_\alpha(u_1, u_2) = \frac{\mathcal{C}_\gamma(du_1, du_2)\mathcal{C}_\gamma(u_1^-, u_2^-)}{\mathcal{C}_\gamma(u_1^-, du_2)\mathcal{C}_\gamma(du_1, u_2^-)}, \quad 0 \leq u_1, u_2 \leq 1, \quad (11)$$

for any pair. Model-checking involves testing the specification of $\theta_\alpha(\cdot, \cdot)$ in (11). This puts the transformed data in the setting considered in §2.3.

The marginal parameters β and $A_{0k}(\cdot)$ may be estimated using a marginal model (Lin, 1994) and the dependence parameter can be estimated using two-stage approaches (Shih & Louis, 1995). Transformed processes $\hat{Y}_{ik}^\alpha(t_k) = I[\exp\{-\hat{A}_{ik}(X_{ik})\} \geq t_k]$ and $\hat{N}_{ik}^\alpha(t_k) = I[\exp\{-\hat{A}_{ik}(X_{ik})\} \leq t_k, \delta_{ik} = 1]$ define the three bivariate hazard functions estimators given in equation (6) with the addition of a superscript α to the notation.

This yields a model-checking process

$$\hat{F}_W^\alpha = \int \int W(u_1, u_2) \left\{ \hat{\Lambda}_{11}^\alpha(du_1, du_2) - \hat{\theta}_\alpha(u_1, u_2) \hat{\Lambda}_{10}^\alpha(du_1, u_2^-) \hat{\Lambda}_{01}^\alpha(u_1^-, du_2) \right\},$$

where integration is over a subset of the unit square. The cross-ratio function can be readily calculated based on estimators of the marginal parameters and the dependence parameter γ . The theory for this process is given in the next section.

4. ASYMPTOTIC THEORY

The asymptotic proofs rely on the theory developed by Gill et al. (1995). The theory is set among right-continuous functions with left-hand limits defined on $[0, \tau_1] \times [0, \tau_2]$, denoted by $D[0, \tau]$ and endowed with the supremum norm, $\|\cdot\|_\infty$. Convergence here implies convergence in supremum norm. Weak convergence is defined in terms of outer probability to avoid the measurability issues inherent in working with a nonseparable space.

The model-checking approach is very broad and can encompass many possible parametric models for the cross-ratio function. Hence, the theory requires some customization for each model: it must be verified that the estimation strategy yields an estimator of the cross-ratio function, $\hat{\theta}(\cdot, \cdot)$, such that, if the frailty distribution is correctly specified, then the following conditions hold,

Condition 1. As $n \rightarrow \infty$, $\hat{\theta} \rightarrow \theta_0$, almost surely.

Condition 2. As $n \rightarrow \infty$, $\sqrt{n}(\hat{\theta} - \theta_0)$ converges weakly to a Gaussian process, \mathcal{Z}_θ , on $D[0, \tau]$.

Condition 3. Given the data, the bootstrapped cross-ratio estimator $\theta^\#(\cdot, \cdot)$ converges in distribution such that $\sqrt{n}(\theta^\# - \hat{\theta})$ converges weakly to \mathcal{Z}_θ almost surely as $n \rightarrow \infty$.

Gill et al. (1995) proved consistency, weak convergence and bootstrap results for the vector of processes $\hat{\Lambda} := (\hat{\Lambda}_{11}, \hat{\Lambda}_{10}, \hat{\Lambda}_{01})$ on the space $(D[0, \tau])^3$. We let \mathcal{Z}_Λ denote the limiting distribution of $\sqrt{n}(\hat{\Lambda} - \Lambda_0)$.

Denote the estimated and true vector of processes by $\hat{\Theta} = (\hat{\Lambda}, \hat{\theta})$ and $\Theta_0 = (\Lambda_0, \theta_0)$. Given the above, it must be further shown that $\sqrt{n}(\hat{\Theta} - \Theta_0)$ converges weakly to \mathcal{Z}_Θ , where \mathcal{Z}_Θ is a Gaussian process on $(D[0, \tau])^4$.

The process \hat{F}_W can be represented as a smooth mapping $\Gamma(\hat{\Theta})$ of the parameter estimators; see (A1) in the Appendix. If the frailty distribution is correctly specified, then $\Gamma(\Theta_0) = 0$ and the model-checking process $\sqrt{n}\hat{F}$ equals $\sqrt{n}\{\Gamma(\hat{\Theta}) - \Gamma(\Theta_0)\}$. It is shown in the Appendix that the map $\Gamma: (D[0, \tau])^4 \rightarrow D[0, \tau]$ is weakly continuously compactly differentiable at Θ_0 , with derivative $d\Gamma_{\Theta_0}$ given as (A2) in the Appendix.

THEOREM 1. *Given assumptions (2)–(4) in Gill et al. (1995), Conditions 1–3, bounded variation of $\hat{\theta}$ and θ , and weak convergence of $\sqrt{n}(\hat{\Theta} - \Theta_0)$ to \mathcal{Z}_Θ , then $\sqrt{n}\hat{F}$ converges weakly to a Gaussian process.*

The result follows immediately from the compact differentiability of Γ and the weak convergence of $\sqrt{n}(\hat{\Theta} - \Theta_0)$. The limiting Gaussian distribution of $\sqrt{n}\hat{F}$ is $d\Gamma(\mathcal{Z}_\Theta) := \mathcal{Z}_F$. Furthermore, $\sqrt{n}\hat{F}$ is asymptotically equivalent to $d\Gamma\{\sqrt{n}(\hat{\Theta} - \Theta_0)\}$.

COROLLARY 1. *Given the assumptions of Theorem 1, for a function W_n on $D[0, \tau]$ with bounded variation and if $W_n \rightarrow W$, which is also bounded and continuous, then $\sqrt{n}\hat{F}_W := \sqrt{n} \int W_n d\hat{F}$ converges weakly to $\int W d\mathcal{Z}_F := \mathcal{Z}_{F_W}$, where \mathcal{Z}_{F_W} is a Gaussian process on $(D[0, \tau])$.*

The proof is given in the Appendix.

Hsu & Prentice (1996b) showed that if $\hat{\theta}(\cdot, \cdot) = 1$ then $\sqrt{n}\hat{F}_W$ forms a test of independence of T_1 and T_2 . They showed that, under independence, $\sqrt{n}\hat{F}_W(t_1, t_2)$ is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^n \int_0^{t_1} \int_0^{t_2} W(u_1, u_2) M_{i1}(du_1) M_{i2}(du_2),$$

where $M_{ik}(t) := N_{ik}(t) - \int_0^t Y_{ik}(u) d\Lambda_k(u)$ is the marginal martingale for T_k with $\Lambda_k(\cdot)$ denoting the k th marginal hazard function for $k = 1, 2$.

The limiting distribution is then Gaussian with independent increments and variance

$$\text{var} \left\{ \sqrt{n}\hat{F}_W(t_1, t_2) \right\} = \int_0^{t_1} \int_0^{t_2} W^2(u_1, u_2) E \{ Y_1(u_1) Y_2(u_2) \} \Lambda_1(du_1) \Lambda_2(du_2).$$

This is also the limiting distribution of a family of independence tests developed by Shih & Louis (1996).

In general, it is difficult to work with the limiting distribution, \mathcal{Z}_{F_W} , analytically; however, it is possible to use the bootstrap for this distribution. This follows because, under Condition 3 and the results of Gill et al. (1995), the bootstrap correctly approximates the limiting distribution of $\sqrt{n}(\hat{\Theta} - \Theta_0)$. Here, compact differentiability is again useful; compactly differentiable transformations preserve the bootstrap (Gill, 1989).

Under the assumptions of Corollary 1 and conditional on the data, the bootstrap process, $\sqrt{n}(F_W^\# - \hat{F}_W)$, converges weakly to \mathcal{Z}_{F_W} almost surely, where $F^\# := \Gamma(\Theta^\#)$. Hence, we can use the bootstrapped process to approximate the null distribution of $\sqrt{n}\hat{F}_W$.

The bootstrap distribution $F_W^\#$ is obtained by sampling with replacement from the vector of processes $(N_{i1}, N_{i2}, Y_{i1}, Y_{i2})$ with substitution into (6) and then into (9). The p -values for E_t and Q can be obtained based on the distributions of

$$E_t^\# = (F_W^\# - \hat{F}_W)(t_1, t_2) / \text{ESE}\{F_W^\#(t_1, t_2)\},$$

$$Q^\# = \sup_{t_1, t_2} |(F_W^\# - \hat{F}_W)(t_1, t_2)|.$$

In the setting of time-fixed covariates, the theory for the model-checking process $\hat{F}_\alpha(\cdot)$ was provided by Fan & Prentice (2002), who constructed detailed proofs that the three hazard functions $(\hat{\Lambda}_{11}^\alpha, \hat{\Lambda}_{10}^\alpha, \hat{\Lambda}_{01}^\alpha)$ are consistent and converge weakly to a Gaussian process whose distribution can be approximated using the bootstrap. Their proofs considered transformation of event times by $\hat{A}(\cdot)$ but the extension to $\exp\{-\hat{A}(\cdot)\}$ is straightforward. The asymptotic theory for model checking with arbitrary cluster sizes using \hat{F}_{1W}^K and \hat{F}_{2W}^K is shown in the Appendix.

Choice of the weight function can be guided by examining the process under a sequence of contiguous alternatives. Suppose the null hypothesis is that the cross-ratio function equals $\theta_0(t_1, t_2)$ with optimal local power desired against a cross-ratio family $\theta_A(t_1, t_2)$. Let $\tilde{\theta}_0(t_1, t_2)$ be the limiting value of $\hat{\theta}_0(t)$ for data under the alternative. Define the series of continuous alternatives H_{An} by the cross-ratio function $\tilde{\theta}_0(t_1, t_2) + n^{-1/2}K(t_1, t_2)$, where $K(t_1, t_2) := \theta_A(t_1, t_2) - \tilde{\theta}_0(t_1, t_2)$. It can be shown that $\sigma(t_1, t_2)^{-1}\sqrt{n}F_W(t_1, t_2)$ converges in distribution to $N\{\kappa(t_1, t_2), 1\}$, where

$$\kappa(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} W(s_1, s_2)K(s_1, s_2)\Lambda_{10}(ds_1, s_2^-)\Lambda_{01}(s_1^-, ds_2)$$

and $\sigma(t_1, t_2) := \text{var}\{\sqrt{n}F_W(t_1, t_2)\}$. By the Cauchy-Schwartz inequality, the noncentrality parameter is maximized when the weight, $W(\cdot, \cdot)$, is proportional to $\theta_A(\cdot, \cdot) - \tilde{\theta}_0(\cdot, \cdot)$. Thus, local power is maximized by selecting weights which are proportional to the differences in the cross-ratio function under the null and alternative models.

5. SIMULATION STUDIES

Simulations examined diagnostics for gamma frailty and positive stable frailty models. For both models, bivariate failure times were generated with exponential margins. We calculated the empirical size and power of 0.05-level tests based on the statistics E_t and Q as a function of cluster size ($n = 50, 100, 200$) and global dependence. Dependence was measured by Kendall's τ , a rank-based measure of intracluster dependence, and was set to 0.25, 0.50 and 0.75. In each setting, the empirical rejection probabilities were calculated using 2000 simulated datasets with critical values based on 200 bootstrap samples.

Example 1: Gamma frailty assessment. The first set of simulations examined the method applied to data from a gamma frailty model when the model was correctly assumed. Here 2000 gamma frailty failure pairs were generated with survivor function equal to

$$S(t_1, t_2; \gamma) = \{\exp(-t_1\gamma) + \exp(-t_2\gamma) - 1\}^{-\gamma^{-1}},$$

with γ values 0.67, 2.0 and 6.0, which corresponded to Kendall's τ of 0.25, 0.50 and 0.75, respectively. The model-checking process was calculated with weight function $\sum Y_{i1}Y_{i2}$. The supremum statistic Q was taken over the grid $[\tau_1, \tau_2] \times [\tau_1, \tau_2]$ such that $S(\tau_1, \tau_1) = 0.90$ and $S(\tau_2, \tau_2) = 0.20$, i.e. 30% censoring, and the time-point for the statistic E_t was (τ_2, τ_2) . Simulation results in Table 1 indicate that, overall, the 0.05-level test has empirical size near the nominal level with a slightly lower rejection rate for highly dependent data and a higher rejection rate for nearly independent data.

To examine the power of the method, paired positive stable datasets were generated with survivor function

$$S(t_1, t_2; \gamma) = \exp\left\{-\left(t_1^{\gamma^{-1}} + t_2^{\gamma^{-1}}\right)^\gamma\right\}, \tag{12}$$

with γ equal to 0.75, 0.50 and 0.25. For the positive stable model, Kendall's $\tau = 1 - \gamma$ and these values correspond to Kendall's τ equal to 0.25, 0.50 and 0.75, respectively. For these datasets, the gamma frailty model was fitted and the test statistics were calculated over the region $[\tau_1, \tau_2] \times [\tau_1, \tau_2]$ with $S(\tau_1, \tau_1) = 0.90$ and $S(\tau_2, \tau_2) = 0.20$.

The results in Table 1 show that the power is similar for the two statistics E_t and Q but varies greatly by scenario. As expected, power increases with the number of pairs and strength of overall dependence. For $\gamma = 0.50$, the current methods have appreciably better power than reported by Glidden (1999). A caveat in the comparison is that the simulations in Glidden (1999) had slightly heavier censoring.

Example 2: Positive stable frailty assessment. Positive stable assessments paralleled the design of the gamma frailty checking. Here we generated positive stable datasets with survivor function (12) for Kendall's τ equal to 0.25, 0.50 and 0.75, for which $\gamma = 0.75, 0.50, 0.25$, respectively. The model-checking process had weight function $W = \sum Y_{i1}Y_{i2}$. The supremum statistic Q was taken over the grid $[\tau_1, \tau_2] \times [\tau_1, \tau_2]$ such that $S(\tau_1, \tau_1) = 0.90$ and $S(\tau_2, \tau_2) = 0.20$, corresponding to 30% censoring. The time-point (τ_2, τ_2) was chosen for E_t .

Simulation results, again in Table 1, indicate that, overall, the 0.05-level test has empirical size slightly higher than the nominal level in most settings. In general, the power for checking the positive stable model appears lower than for checking the gamma frailty model, even with comparable dependence. Finally, the supremum statistic Q has consistently more power than the E_t statistic.

6. DATA EXAMPLE

The Australian Twins Study (Duffy et al., 1990) analyzed the association between monozygotic twins and dizygotic twins in various diseases. The study collected information on a number of diseases, including appendicitis. Subjects were asked if they had undergone

Table 1. *Empirical sizes/powers of the proposed tests. Two null hypotheses considered are that data follow the gamma frailty model or the positive stable frailty model*

True Model	Kendall's τ	Statistic	Sample size			Sample size		
			50	100	200	50	100	200
			H_0 : Gamma			H_0 : Positive Stable		
Gamma frailty	0.25	E_τ	0.06	0.07	0.07	0.38	0.66	0.92
		Q	0.05	0.06	0.07	0.20	0.49	0.83
	0.50	E_τ	0.06	0.06	0.05	0.88	1.00	1.00
		Q	0.04	0.04	0.05	0.66	0.97	1.00
	0.75	E_τ	0.03	0.05	0.04	0.97	1.00	1.00
		Q	0.02	0.03	0.04	0.85	1.00	1.00
Positive stable	0.25	E_τ	0.34	0.57	0.84	0.08	0.07	0.06
		Q	0.35	0.60	0.89	0.06	0.04	0.06
	0.50	E_τ	0.74	0.97	1.00	0.07	0.07	0.06
		Q	0.73	0.96	1.00	0.04	0.05	0.06
	0.75	E_τ	0.81	1.00	1.00	0.09	0.06	0.06
		Q	0.70	1.00	1.00	0.03	0.06	0.05

appendectomy and, if so, at what age the procedure was performed. We apply our methods to female monozygotic twin pairs in the study; the dataset comprises 1214 twin pairs with 590 appendectomies. To assess the effect of weight functions, $W_n(\cdot, \cdot)$, we compared results for a unit weight function and a weight function equal to the size of the bivariate risk set, $\sum Y_{i1}(\cdot)Y_{i2}(\cdot)$.

The fit of the gamma frailty model to the monozygotic twin pairs yields an estimated cross-ratio function of $\hat{\theta}(t_1, t_2) = 2.84$ for all (t_1, t_2) , with 95% confidence interval from 2.36 to 3.41. Previous analysis by Fan, Hsu & Prentice (2000) suggested that the cross-ratio does not appear to be constant; instead, it appears to be decreasing in time. Glidden (1999) found suggestive, but nonsignificant, departure from a gamma frailty model in this dataset.

The cumulative residual plots assuming the correctness of the gamma frailty model appear in Fig. 2(a). In addition, we present the process $\hat{F}_W(t, t)$, $t \in [0, 60]$, in Fig. 2(b) along with 25 processes sampled from the bootstrap approximation to the null distribution. The gamma frailty model is unequivocally rejected for these data. The p -value from the 10 000 bootstrap realizations from the null distribution using the risk set weight is 0.0026. Figures 2(a) and (b) suggest that the model of a constant cross-ratio function for the gamma frailty tends to underestimate the cross-ratio function prior to age 20. Around this age, the cross-ratio function diminishes, a pattern found in positive stable models. The results were similarly significant for the unit weight, which gave a p -value of 0.0017. Fitting the positive stable model gives a maximum pseudolikelihood estimate of $\hat{\gamma} = 0.76$ with 95% bootstrap percentile confidence interval (0.71, 0.80). Model-checking for the positive stable model yielded a p -value of 0.38 from 10 000 bootstrap realizations for the risk-set weight and 0.13 for the unit weight, suggesting reasonable agreement with the positive stable model.

The Frank's family (Genest, 1987) with copula

$$C_\gamma(u_1, u_2) = \log_\gamma \left\{ 1 + \frac{(\gamma^{u_1} - 1)(\gamma^{u_2} - 1)}{\gamma - 1} \right\}$$

was also fitted to the data. Like the positive stable model, the above copula yields a decreasing cross-ratio function with time but it does not decay to

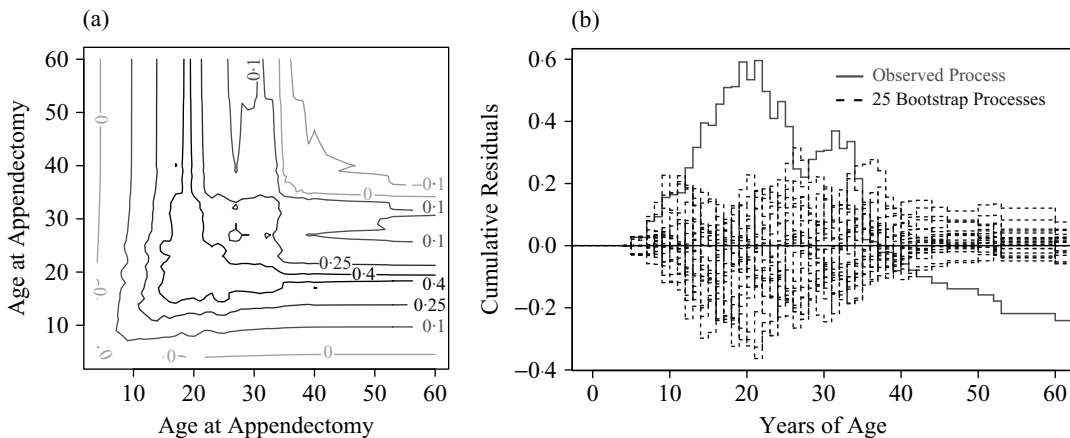


Fig. 2. Australian monozygotic twins data. (a) Contour plot of residual process \hat{F}_{2W}^K for the correctness of the gamma frailty model. (b) The process $\hat{F}_W(t, t)$ from $t = 0$ to 60 years of age along with 25 bootstrap processes sampled from the null distribution.

local independence as steeply as the positive stable model. The maximum pseudolikelihood estimate of γ is 0.031 for the twins data, which translates to a global Kendall's τ of 0.35. The model-checking method finds poor agreement of the Frank's copula with the data. From 10 000 bootstrap realizations, we obtain p -values of 0.032 and 0.018 using the unit and risk-set weights, respectively.

7. DISCUSSION

The diagnostics developed in this paper can, in principle, check any frailty or copula model for multivariate failure-time data. Since checking is based on examining the cross-ratio function, it is a natural accompaniment to the methods of Fan, Hsu & Prentice (2000) and Fan, Prentice & Hsu (2000) for exploring the shape of the cross-ratio function. The approach has the desirable feature that the bootstrap can be used as the basis for inference, but care is needed to verify that a particular copula estimation strategy satisfies the conditions in §4. A possible extension of the method could use smoothed rather than cumulative sums of residuals.

Simulations seem to indicate that this method outperforms the method of Glidden (1999). The method of Glidden (1999) relies on aligning failures within and across clusters on to a common time scale. However, the cross-ratio function is naturally a function of bivariate time and it appears that both interpretation and power are lost by forcing model-checking on to a single time dimension.

Finally, the data setting considered in this paper is one in which families are sampled at random from the population. In practice, pedigrees are often ascertained based on diseased probands. The extension of the method to ascertained samples is compelling but not immediate.

ACKNOWLEDGEMENT

This research was supported by the U.S. National Heart, Lung and Blood Institute. The author thanks Charles E. McCulloch, the editor, the associate editor and a reviewer for their helpful advice and comments.

APPENDIX

Technical details

Let $\Theta = (\Lambda_{11}, \Lambda_{10}, \Lambda_{01}, \theta)$ denote the full parameter vector with the estimated and underlying values denoted by $\hat{\Theta}$ and Θ_0 , respectively.

The metric on this space is the supremum norm, denoted by $\|\cdot\|_\infty$. The variation norm, denoted by $\|\cdot\|_v$ is defined as in Gill et al. (1995). We repeatedly use their Lemmas 2.2, integration by parts, and 2.5, Helly-Bray, and closely follow the general approach of their proofs in §5.

Compact differentiability of Γ . We show here that the mapping Γ ,

$$\Gamma(\Theta) := \int \Lambda_{11}(dt_1, dt_2) - \int \theta(t_1, t_2) \Lambda_{10}(dt_1, t_2^-) \Lambda_{01}(t_1^-, dt_2), \quad (\text{A1})$$

is a weakly continuously compactly differentiable map tangential to a function of bounded variation.

The map is from $(D[0, \tau])^4$ to $(D[0, \tau])$ with a derivative, denoted by $d\Gamma$, at Θ_0 which maps $h = (h_{11}, h_{10}, h_{01}, h_\theta)$ to

$$\begin{aligned} d\Gamma_{\Theta_0}(h) = & \int h_{11}(dt_1, dt_2) - \int h_\theta(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) \\ & - \int \theta(t_1, t_2)h_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) \\ & - \int \theta(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)h_{01}(t_1^-, dt_2), \end{aligned} \quad (\text{A2})$$

which is also a map from $(D[0, \tau])^4$ to $D[0, \tau]$.

To show that Γ is compactly differentiable, it must be shown that, for a sequence $h^n = \sqrt{n}(\hat{\Theta} - \Theta_0) \rightarrow h$,

$$|\sqrt{n}\{\Gamma(\hat{\Theta}) - \Gamma(\Theta_0)\} - d\Gamma_{\Theta_0}(h^n)|_\infty \rightarrow 0, \quad (\text{A3})$$

and this we now do.

The difference $\sqrt{n}\{\Gamma(\hat{\Theta}) - \Gamma(\Theta_0)\}$ equals

$$\begin{aligned} & \int h_{11}^n(dt_1, dt_2) - \int h_\theta^n(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) \\ & - \int \hat{\theta}(t_1, t_2)h_{10}^n(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) \\ & - \int \hat{\theta}(t_1, t_2)\hat{\Lambda}_{10}(dt_1, t_2^-)h_{01}^n(t_1^-, dt_2). \end{aligned} \quad (\text{A4})$$

The difference in (A3) is just the difference between (A4) and (A2), which equals

$$- \int (h_\theta^n - h_\theta)(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)\Lambda_{01}(t_1^-, dt_2) \quad (\text{A5})$$

$$- \int \left\{ \hat{\theta}(t_1, t_2)h_{10}^n(dt_1, t_2^-) - \theta(t_1, t_2)h_{10}(dt_1, t_2^-) \right\} \Lambda_{01}(t_1^-, dt_2) \quad (\text{A6})$$

$$- \left\{ \int \hat{\theta}(t_1, t_2)\hat{\Lambda}_{10}(dt_1, t_2^-)h_{01}^n(t_1^-, dt_2) - \int \theta(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)h_{01}(t_1^-, dt_2) \right\}. \quad (\text{A7})$$

We show that (A7) converges to zero almost surely. Similar approaches can show that (A5) and (A6) converge to zero as well.

The term (A7) can be written as

$$- \int \hat{\theta}(t_1, t_2)\hat{\Lambda}_{10}(dt_1, t_2^-)(h_{01}^n - h_{01})(t_1^-, dt_2) \quad (\text{A8})$$

$$- \int \hat{\theta}(t_1, t_2)(\hat{\Lambda}_{10} - \Lambda_{10})(dt_1, t_2^-)h_{01}(t_1^-, dt_2) \quad (\text{A9})$$

$$- \int (\hat{\theta} - \theta)(t_1, t_2)\Lambda_{10}(dt_1, t_2^-)h_{01}(t_1^-, dt_2). \quad (\text{A10})$$

Convergence of (A9) and (A10) is complicated by the unbounded variation of h_{01} . The approach here is to use the Helly-Bray lemma as Gill et al. (1995) did in their proof of Proposition 3.2. The function h_{01} can be approximated to arbitrary tolerance by a sequence of functions h_{01}^m , $m \rightarrow \infty$, with bounded variation which diverges with m . We then replace h_{01} with $(h_{01} - h_{01}^m) + h_{01}^m$. The term with $(h_{01} - h_{01}^m)$ can be bounded by the sup norm of $(h_{01} - h_{01}^m)$ times a constant while the term

with h_{01}^m is bounded above by the sup norm of $\hat{\Lambda}_{10} - \Lambda_{10}$ times the variation of h_{01}^m . If we allow m to diverge to ∞ slowly enough, the term converges. A similar proof applies for (A8) and (A9) and thus Γ is a weakly continuous compact differentiable map.

Weak convergence of \hat{F}_W . Given the compact differentiability of Γ , the weak convergence of $\sqrt{n}\hat{F}$ is shown in §4. Next, it remains to be proved that if $W_n \rightarrow W$ and $\sqrt{n}\hat{F}_W \Rightarrow \mathcal{F}_W$ then $\sqrt{n} \int W_n d\hat{F} \Rightarrow \int W dF$. This is a Slutsky-type result. Using a Skorohod construction, consider a probability space where $\sqrt{n}\hat{F} \rightarrow \mathcal{F}$. Since all functions have bounded variation, then $\sqrt{n} \int W_n d\hat{F} \rightarrow \int W dF := \mathcal{F}_W$; this proves weak convergence on the original probability space.

Asymptotic theory for \hat{F}_{1W}^K and \hat{F}_{2W}^K . Let $\mathcal{H}(t_1, \dots, t_K; s_1, \dots, s_K)$ be a process equal to

$$\text{pr}(T_1 \geq t_1, \dots, T_K \geq t_K, C_1 \geq s_1, \dots, C_K \geq s_K).$$

Define $H_{kl}(t_1, t_2; s_1, s_2) = \text{pr}(T_k \geq t_1, T_l \geq t_2, C_k \geq s_1, C_l \geq s_2)$, which is a simple linear mapping from \mathcal{H} obtained by setting $t_m = s_m = 0$ for all $m \neq k, m \neq l$.

The extended hazard functions in §3.1 can be written as

$$\begin{aligned} \Lambda_{11}^K(t_1, t_2) &= - \int_0^{t_1} \int_0^{t_2} \frac{\sum_{k \neq l} H_{kl}(ds_1, ds_2; s_1, s_2)}{\sum_{k \neq l} H_{kl}(s_1, s_2; s_1, s_2)} \\ \Lambda_{01}^K(t_1, t_2) &= - \int_0^{t_2} \frac{\sum_{k \neq l} H_{kl}(t_1, ds_2, t_1, s_2)}{\sum_{k \neq l} H_{kl}(t_1, s_2; t_1, s_2)} \\ \Lambda_{10}^K(t_1, t_2) &= - \int_0^{t_1} \frac{\sum_{k \neq l} H_{kl}(ds_1, t_2; s_1, t_2)}{\sum_{k \neq l} H_{kl}(s_1, t_2; s_1, t_2)}. \end{aligned}$$

It is straightforward to show that

$$(\Lambda_{11}^K, \Lambda_{10}^K, \Lambda_{01}^K) = \phi_1(\mathcal{H}),$$

where ϕ_1 is the composition of a series of compactly differentiable maps, cf. division and integration, and thus by the chain rule ϕ_1 is compactly differentiable. Likewise, if

$$\mathcal{H}_n(t_1, \dots, t_K; s_1, \dots, s_K) = n^{-1} \sum_{i=1}^n I(T_{i1} \geq t_1, \dots, T_{iK} \geq t_K, C_{i1} \geq s_1, \dots, C_{iK} \geq s_K)$$

then

$$(\hat{\Lambda}_{11}^K, \hat{\Lambda}_{10}^K, \hat{\Lambda}_{01}^K) = \phi_1(\mathcal{H}_n).$$

We also defined the pairwise hazard functions as

$$\begin{aligned} \Lambda_{11}^{kl}(t_1, t_2) &= - \int_0^{t_1} \int_0^{t_2} \frac{H_{kl}(ds_1, ds_2; s_1, s_2)}{H_{kl}(s_1, s_2; s_1, s_2)} \\ \Lambda_{10}^{kl}(t_1, t_2) &= - \int_0^{t_2} \frac{H_{kl}(t_1, ds_2, t_1, s_2)}{H_{kl}(t_1, s_2; t_1, s_2)} \\ \Lambda_{01}^{kl}(t_1, t_2) &= - \int_0^{t_1} \frac{H_{kl}(ds_1, t_2; s_1, t_2)}{H_{kl}(s_1, t_2; s_1, t_2)}. \end{aligned}$$

This vector of functions can be written in the form $(\Lambda_{11}^{kl}, \Lambda_{10}^{kl}, \Lambda_{01}^{kl}, k \neq l)$ equals $\phi_2(\mathcal{H})$, where ϕ_2 is also a compactly differentiable mapping. Here the empirical hazard functions $(\hat{\Lambda}_{11}^{kl}, \hat{\Lambda}_{10}^{kl}, \hat{\Lambda}_{01}^{kl}, k \neq l)$ are equal to $\phi_2(\mathcal{H}_n)$.

Since both sets of hazard functions can be written as compactly differentiable maps of empiricals, a functional delta method for the maps $\phi_1(\cdot)$, $\phi_2(\cdot)$ and $\Gamma(\cdot)$ combined with empirical process theory for the process $H(\cdot)$ establishes weak convergence of $\sqrt{n}\hat{F}_{1W}^K$ and $\sqrt{n}\hat{F}_{2W}^K$ to Gaussian processes.

REFERENCES

- CHEN, M. C. & BANDEEN-ROCHE, K. (2005). A diagnostic for association in bivariate survival models. *Lifetime Data Anal.* **11**, 245–64.
- 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**, 141–51.
- DABROWSKA, D. M. (1988). Kaplan-Meier estimators on the plane. *Ann. Statist.* **16**, 1475–89.
- DUFFY, D. L., MARTIN, N. G. & MATHEWS, J. D. (1990). Appendectomy in Australian twins. *Am. J. Hum. Genet.* **47**, 590–2.
- FAN, J. J., HSU, L. & PRENTICE, R. L. (2000). Dependence estimation over a finite bivariate failure time region. *Lifetime Data Anal.* **6**, 343–55.
- FAN, J. J., PRENTICE, R. L. & HSU, L. (2000). A class of weighted dependence measures for bivariate failure time data. *J. R. Statist. Soc. B* **62**, 181–90.
- FAN, J. J. & PRENTICE, R. L. (2002). Covariate-adjusted dependence estimation on a finite bivariate failure time region. *Statist. Sinica* **12**, 689–705.
- GENEST, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549–55.
- GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). *Scand. J. Statist.* **16**, 97–128.
- GILL, R. D., VAN DER LAAN, M. J. & WELLNER, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincaré* **31**, 545–97.
- GLIDDEN, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika* **86**, 381–93.
- HOUGAARD, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671–8.
- HSU, L. & PRENTICE, R. L. (1996a). On assessing the strength of dependency between failure time variates. *Biometrika* **83**, 491–506.
- HSU, L. & PRENTICE, R. L. (1996b). A generalisation of the Mantel-Haenszel test to bivariate failure time data. *Biometrika* **83**, 905–11.
- LIN, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Med.* **13**, 2233–47.
- NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. & SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–43.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.
- PRENTICE, R. L. & CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512.
- SHIH, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika* **85**, 189–200.
- SHIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384–99.
- SHIH, J. H. & LOUIS, T. A. (1996). Tests of independence for bivariate survival data. *Biometrics* **52**, 1440–9.
- VISWANATHAN, B. & MANATUNGA, A. K. (2001). Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Anal.* **7**, 143–55.

[Received August 2005. Revised August 2006]