J. R. Statist. Soc. B (2003)
65, *Part* 1, *pp.* 257–273

Estimating the association parameter for copula models under dependent censoring

Weijing Wang

National Chiao-Tung University, Hsin-Chu, Taiwan

[Received August 2000. Revised August 2002]

Summary. Many biomedical studies involve the analysis of multiple events. The dependence between the times to these end points is often of scientific interest. We investigate a situation when one end point is subject to censoring by the other. The model assumptions of Day and co-workers and Fine and co-workers are extended to more general structures where the level of association may vary with time. Two types of estimating function are proposed. Asymptotic properties of the proposed estimators are derived. Their finite sample performance is studied via simulations. The inference procedures are applied to two real data sets for illustration.

Keywords: Archimedean copula models; Bivariate survival analysis; Competing risk; Cross-ratio function; Estimating function; Frailty models; Identifiability; Kendall's τ; Log-rank statistic; Multistate process; Semi-competing-risks data; Semiparametric inference

1. Introduction

There has been substantial scientific interest in studying the dependence between two failure time variables. Many statistical methods have been proposed under the context of bivariate survival analysis. In this paper, we focus on a special situation when one variable is subject to censoring by the other, something that often occurs in data with multiple end points. Let (T_1, T_2) be two failure time variables and suppose that their relationship in the region of $T_1 \leqslant T_2$ is of main interest. For example, in a study of bone marrow transplants, T_1 denotes the time to relapse of leukaemia and T_2 is the time to death. The relationship between T_1 and T_2 has important biological meaning. Consider another example in research into acquired immune deficiency syndrome (AIDS), where T_1 is defined as the time when CD4 cell counts decrease to some critical level and T_2 as the time to occurrence of AIDS. For predicting AIDS, only the joint relationship before the occurrence of AIDS is relevant. High correlation between T_1 and T_2 given $T_1 \leqslant T_2$ implies good predictability of CD4 cell counts as a marker for AIDS. In both cases, T_1 is subject to censoring by T_2. Such a special dependent censoring structure complicates statistical inference.

Let C be an external censoring variable due to withdrawal of patients or the end of study. We observe the variables $X = T_1 \wedge T_2 \wedge C$, $\delta_1 = I\{T_1 \leqslant (T_2 \wedge C)\}$, $Y = T_2 \wedge C$ and $\delta_2 = I(T_2 \leqslant C)$, where '$\wedge$' denotes the minimum and $I(\cdot)$ is the indicator function. $T_1 \wedge T_2$ is also subject to censoring by C with the indicator, $\delta_0 = I(T_1 \wedge T_2 \leqslant C) = \delta_1 + \delta_2 - \delta_1\delta_2$. Let $\{(T_{1i}, T_{2i}, C_i)$ $(i = 1, \ldots, n)\}$ be independent and identically distributed (IID) replications of (T_1, T_2, C). The observed data are IID replications of $(X, Y, \delta_1, \delta_2, \delta_0)$:

Address for correspondence: Weijing Wang, Institute of Statistics, National Chiao-Tung University, Hsin-Chu, Taiwan, Republic of China.
E-mail: wjwang@stat.nctu.edu.tw

 1369–7412/03/65257

$$\{(X_i, Y_i, \delta_{1i}, \delta_{2i}, \delta_{0i})\ (i = 1, \ldots, n)\}. \quad (1)$$

The data structure in expression (1) is called 'semi-competing-risks data' by Fine *et al.* (2001) because T_2 is a competing risk for T_1 but not vice versa. Let $F_k(t) = \text{pr}(T_k \geqslant t)$ and $G(t) = \text{pr}(C \geqslant t)$ be the marginal survival functions of T_k $(k = 1, 2)$ and C respectively. Also let $F(s,t) = \text{pr}(T_1 \geqslant s, T_2 \geqslant t)$ be the joint survival function of (T_1, T_2) and $F_b(t) = \text{Pr}(T_1 \wedge T_2 \geqslant t)$ be the survival function of $T_1 \wedge T_2$. It is important to mention that $F_1(\cdot)$ is not identifiable because T_1 is not observable in $T_1 > T_2$. Given data (1), the Kaplan–Meier method can be applied to estimate $F_b(t)$, $F_2(t)$ and $G(t)$ nonparametrically. $G(t)$ can be estimated by the Kaplan–Meier estimator either on the basis of $\{(X_i, 1-\delta_{0i})\ (i = 1, \ldots, n)\}$ or $\{(Y_i, 1-\delta_{2i})\ (i = 1, \ldots, n)\}$. For $s \leqslant t$, $F(s,t)$ can be estimated by

$$\hat{F}(s,t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t)/n\,\hat{G}(t),$$

where $\hat{G}(t)$ is an estimator of $G(t)$.

The major goal of this paper is to study the dependence relationship between (T_1, T_2) in the upper wedge, $\mathcal{P} = \{(s,t) : 0 < s \leqslant t < \infty\}$. In principle, we can examine whether

$$F(s,t)/F_1(s)\,F_2(t) = 1$$

to assess the existence of local dependence at bivariate time (s,t). However $F_1(s)$ is not identifiable nonparametrically, which implies that such an assessment requires extra assumptions. Two other references address the same problem as the present paper. Day *et al.* (1997) considered the predictive hazard ratio function

$$\theta(s,t) = \frac{\lambda_2(t|T_1 = s)}{\lambda_2(t|T_1 > s)} \qquad (s,t) \in \mathcal{P}, \quad (2)$$

where $\lambda_2(t|A)$ is the hazard function of T_2 given that event A occurs. Assuming that $\theta(s,t) = \alpha$ for $(s,t) \in \mathcal{P}$, Day *et al.* (1997) proposed an estimating function for α based on data (1). It can be shown that, if (T_1, T_2) follow the Clayton model with

$$F(s,t) = \{F_1(s)^{1-\alpha} + F_2(t)^{1-\alpha}\}^{1/(1-\alpha)} \qquad (s,t) \in \mathcal{P}, \quad (3)$$

then $\theta(s,t) = \alpha$ for $(s,t) \in \mathcal{P}$. The converse relationship is also true for Clayton's model defined on $[0,\infty)^2$. Alternatively, Fine *et al.* (2001) defined the concordance indicator, $\Delta_{ij} = I\{(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0\}$ $(i \neq j)$, where (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}) are IID replications of (T_1, T_2). They showed that, if model (3) is true,

$$E(\Delta_{ij}) = E(\Delta_{ij}|\tilde{X}_{ij} < \tilde{Y}_{ij}) = \frac{\alpha}{1+\alpha}, \quad (4)$$

where $\tilde{X}_{ij} = T_{1i} \wedge T_{1j}$ and $\tilde{Y}_{ij} = T_{2i} \wedge T_{2j}$. Under model (4) Fine *et al.* (2001) proposed an estimating function for α in the form of a U-statistic.

The Clayton model is a special case of a general copula model with the joint survival function

$$F(s,t) = \text{pr}(T_1 > s, T_2 > t) = C_\alpha\{F_1(s), F_2(t)\}, \quad (5)$$

where $C_\alpha(u,v) : [0,1]^2 \to [0,1]$. Copula models have the desirable feature that the dependence structure is modelled separately from the marginal distributions. The parameter α measures global association and is related to Kendall's τ via

$$\tau = 4\int_0^1\int_0^1 C_\alpha(u,v)\,C_\alpha(\mathrm{d}u, \mathrm{d}v) - 1, \quad (6)$$

where τ denotes Kendall's τ defined as

$$\tau = \Pr\{(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0\} - \Pr\{(T_{1i} - T_{1j})(T_{2i} - T_{2j}) < 0\}. \tag{7}$$

The Archimedean copula (AC) family is a useful subclass of copula models with

$$F(s, t) = \phi_\alpha^{-1}[\phi_\alpha\{F_1(s)\} + \phi_\alpha\{F_2(t)\}], \tag{8}$$

where $\phi(\cdot)$ is a decreasing convex function defined on (0, 1] satisfying $\phi(1) = 0$. The bivariate frailty family discussed in Oakes (1989) belongs to the AC family such that $\phi^{-1}(\cdot)$ is the Laplace transform of the underlying frailty distribution.

Let us consider two general dependence structures defined on the upper wedge $\mathcal{P}$.

(a) Model 1: $\theta(s, t)$ in equation (2) can be further parameterized as $\theta_{\alpha,\eta}(s, t)$ for $(s, t) \in \mathcal{P}$, where α is a one-dimensional parameter of interest and η denotes the nuisance parameter.
(b) Model 2: (T_1, T_2) jointly follow a copula model in equation (5) for $(s, t) \in \mathcal{P}$.

Under special cases, these two assumptions describe dual relationships. For example Oakes (1989) showed that, for an (unrestricted) AC model (8), $\theta_{\alpha,\eta}(s, t) = \tilde{\theta}_\alpha\{F(s, t)\}$, where $\eta = F(s, t)$ and

$$\tilde{\theta}_\alpha(v) = -v \frac{\partial^2\phi_\alpha(v)/\partial v^2}{\partial\phi_\alpha(v)/\partial v} = -v \frac{\phi''_\alpha(v)}{\phi'_\alpha(v)}. \tag{9}$$

However, the relationship between the two models becomes less clear when they are imposed only for the upper wedge.

The underlying model assumption affects subsequent inference procedures. Two types of estimating equation based on model 1 are presented in Section 2. Inference procedures based on model 2 are discussed in Section 3. In Section 4.1 we examine the finite sample performance of the estimators via simulations. The methods proposed are applied to two real examples in Section 4.2 and some concluding remarks are given in Section 5.

2. Inference procedures based on model 1

The methods presented in this section generalize the work by Day *et al.* (1997).

2.1. Derivation from the log-rank statistic

Given observed bivariate failure times (s, t), we can construct the 2×2 table (Table 1) with margins '$T_1 = s$' *versus* '$T_1 > s$' and '$T_2 = t$' *versus* '$T_2 > t$'. The notation in each cell denotes the observed cell count, where

$$N_{11}(\mathrm{d}s, \mathrm{d}t) = \sum_{i=1}^{n} I(X_i = s, \delta_{1i} = 1, Y_i = t, \delta_{2i} = 1),$$

Table 1. 2 × 2 table

	$T_2 = t$	$T_2 > t$	
$T_1 = s$	$N_{11}(\mathrm{d}s, \mathrm{d}t)$		$N_{10}(\mathrm{d}s, t)$
$T_1 > s$			
	$N_{01}(s, \mathrm{d}t)$		$R(s, t)$

$$N_{10}(\mathrm{d}s, t) = \sum_{i=1}^{n} I(X_i = s, \delta_{1i} = 1, Y_i \geqslant t),$$

$$N_{01}(s, \mathrm{d}t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i = t, \delta_{2i} = 1),$$

$$R(s, t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t).$$

Conditioning on the marginal counts, $N_{11}(\mathrm{d}s, \mathrm{d}t)$ follows a hypergeometric distribution with mean

$$\tilde{E}_{11}(\mathrm{d}s, \mathrm{d}t; \alpha, \eta) = \frac{\theta_{\alpha,\eta}(s, t)\, N_{10}(\mathrm{d}s, t)\, N_{01}(s, \mathrm{d}t)}{\theta_{\alpha,\eta}(s, t)\, N_{10}(\mathrm{d}s, t) + R(s, t) - N_{10}(\mathrm{d}s, t)}. \tag{10}$$

Plugging in $\hat{\eta}$, an estimator of η, and summing over the grid formed by the observed failure points in $\mathcal{P}$, we obtain the estimating function

$$L(\alpha, \hat{\eta}) = n^{-1} \int\int_{(s,t)\in\mathcal{P}} w(s, t)\{N_{11}(\mathrm{d}s, \mathrm{d}t) - \tilde{E}_{11}(\mathrm{d}s, \mathrm{d}t; \alpha, \hat{\eta})\}, \tag{11}$$

where $w(s, t)$ is a weight function. A solution to $L(\alpha, \hat{\eta}) = 0$ yields an estimator of α, denoted as $\hat{\alpha}$. Numerically, $\hat{\alpha}$ can be obtained by iterating the equation

$$\hat{\alpha}^{(k+1)} = \hat{\alpha}^{(k)} - \left\{ \left.\frac{\partial L(\alpha, \hat{\eta})}{\partial \alpha}\right|_{\hat{\alpha}^{(k)}} \right\}^{-1} L(\hat{\alpha}^{(k)}, \hat{\eta}),$$

where $\hat{\alpha}^{(k)}$ is the estimated value of α at the kth iteration and $\hat{\alpha}^{(0)}$ is the initial value. An implicit assumption of this procedure is that η can be directly estimated by $\hat{\eta}$ or indirectly by $\hat{\eta}(\alpha)$ given the value of α. When $\theta(s, t) = \alpha$ and $L(\alpha, \hat{\eta}) = 0$, $L(\alpha)$ reduces to the estimating function that was proposed by Day *et al.* (1997). When T_1 and T_2 are independent in the upper wedge, $\alpha = 1$ and thus $L(1)$ can be used for testing the hypothesis of independence. The test procedure proposed by Hsu and Prentice (1996) used a similar idea under independent right censoring.

2.2. Derivation from Doob–Meyer decomposition

The Doob–Meyer decomposition of a counting process can be used to construct estimating functions (Fleming and Harrington (1991), section 2.2). Consider the filtration

$$\mathcal{F}^*_{s,t} = \sigma\{I(X_i \leqslant u, \delta_{1i} = 1), I(Y_i \leqslant v, \delta_{2i} = 1), u \leqslant s, v \leqslant t,\ i = 1, 2, \ldots, n\}, \tag{12}$$

which describes the history of the non-terminal process up to time s and that of the terminal event process up to time t. For a fixed value s, denote $\mathcal{F}^*_s = \{\mathcal{F}^*_{s,t} : t \in (s, \infty)\}$. The compensator of $I(Y_i \leqslant t, \delta_{2i} = 1)$ with respect to $\mathcal{F}^*_s$ can be derived as

$$v_i(t; s) = \int_s^t E\{\mathrm{d}I(Y_i \leqslant u, \delta_{2i} = 1)|\mathcal{F}_{s,u-}\}, \tag{13}$$

where

$$\mathrm{d}v_i(t; s) = I(Y_i \geqslant t)\bigg[\{1 - I(X_i \leqslant s, \delta_{1i} = 1)\}\, \lambda_2(t|T_1 > s) + \int_0^s \lambda_2(u|T_1 = w)\, \mathrm{d}I(X_i \leqslant w, \delta_{1i} = 1)\bigg].$$

It follows that

$$E\{\mathrm{d}I(X_i = s, \delta_{1i} = 1, Y_i \leqslant t, \delta_{2i} = 1)|\mathcal{F}_{s,u-}\} = I(X_i = s, \delta_{1i} = 1)\, I(Y_i \geqslant t)\, \Lambda_2(\mathrm{d}t|T_1 = s)$$
$$= I(X_i = s, \delta_{1i} = 1, Y_i \geqslant t)\, \theta_{\alpha_0,\eta_0}(s,t)\, \Lambda_2(\mathrm{d}t|T_1 > s),$$

where α_0 and η_0 are the true values of α and η respectively. Combining all the observed failure points in $\mathcal{P}$, we obtain the estimating function

$$\sum_{i=1}^{n} \int\int_{(s,t)\in\mathcal{P}} w(s,t)\, I(X_i = s, \delta_{1i} = 1)\{\mathrm{d}I(Y_i \leqslant t, \delta_{2i} = 1) - I(Y_i \geqslant t)\, \theta_{\alpha,\eta}(s,t)\, \Lambda_2(\mathrm{d}t|T_1 > s)\}$$
$$= \int\int_{(s,t)\in\mathcal{P}} w(s,t)\, \{N_{11}(\mathrm{d}s, \mathrm{d}t) - \theta_{\alpha,\eta}(s,t)\, N_{10}(\mathrm{d}s,t)\, \Lambda_2(\mathrm{d}t|T_1 > s)\},$$

where the weight function $w(s,t)$ is predictable with respect to $\mathcal{F}^*_{s,t}$.

Note that $\Lambda_2(t|T_1 > s)$ is also a nuisance parameter which can be estimated by

$$\hat{\Lambda}_2(t|T_1 > s; \alpha, \eta) = \int_{v\leqslant t} \frac{N_{01}(s, \mathrm{d}v)}{\theta_{\alpha,\eta}(s,v)\, N_{10}(\mathrm{d}s, v) + R(s,v) - N_{10}(\mathrm{d}s, v)}, \qquad (14)$$

or by

$$\tilde{\Lambda}_2(t|T_1 > s) = \int_{v\leqslant t} \frac{N_{01}(s+, \mathrm{d}v)}{R(s+, v)}. \qquad (15)$$

Using the plug-in approach, we can construct the estimating functions

$$L_1(\alpha, \hat{\eta}) = n^{-1} \int_{0<s<\infty} \int_{s\leqslant t} w(s,t)\{N_{11}(\mathrm{d}s, \mathrm{d}t) - \theta_{\alpha,\hat{\eta}}(s,t)\, N_{10}(\mathrm{d}s,t)\, \hat{\Lambda}_2(\mathrm{d}t|T_1 > s, \alpha, \hat{\eta})\}, \qquad (16)$$

and

$$L_2(\alpha, \hat{\eta}) = n^{-1} \int_{0<s<\infty} \int_{s\leqslant t} w(s,t)\{N_{11}(\mathrm{d}s, \mathrm{d}t) - \theta_{\alpha,\hat{\eta}}(s,t)\, N_{10}(\mathrm{d}s,t)\, \tilde{\Lambda}_2(\mathrm{d}t|T_1 > s)\}. \qquad (17)$$

α can be estimated by $\hat{\alpha}_j$, the solution to $L_j(\alpha, \hat{\eta}) = 0$ $(j = 1, 2)$. It is easy to see that $L_1(\alpha, \hat{\eta})$ equals $L(\alpha, \hat{\eta})$ given in equation (11).

$\hat{\Lambda}_2(t|T_1 > s, \alpha, \eta)$ in equation (14) seems less intuitive and more complicated than $\tilde{\Lambda}_2(t|T_1 > s)$ in equation (15). In Appendix A, we show that these two estimators are identical with probability 1 as $n \to \infty$. In Appendix B, it is proved that $L_1(\alpha, \eta_0)$ yields an unbiased estimating equation whereas $L_2(\alpha, \eta_0)$ does not. However, when η_0 needs to be estimated, $L_1(\alpha, \hat{\eta})$ does not have an obvious advantage. The finite sample performance of the estimators under different situations will be compared in Section 4.

2.3. Large sample properties

To simplify the analysis, we assume that $\hat{\eta}$ is independent of α. If $\hat{\eta} = \hat{\eta}(\alpha)$, $L_j\{\alpha, \hat{\eta}(\alpha)\} = 0$ is still a valid estimating equation if it produces a unique root. The following regularity conditions will be used for establishing the asymptotic results.

(a) *Condition 1*: $w(s,t)$ is predictable with respect to $\mathcal{F}^*_{s,t}$, bounded and of bounded variation; $\theta_{\alpha,\eta}(s,t)$ is a monotone function of α and twice differentiable with respect to α with bounded derivatives; $\theta_{\alpha,\eta}(s,t)$ is a continuous function of η and $\hat{\eta} \to^{\mathrm{p}} \eta_0$.
(b) *Condition 2*: $\theta_{\alpha,\eta}(s,t)$ is twice differentiable with respect to η and the derivatives are bounded for all $(s,t) \in \mathcal{P}$ and $n^{1/2}(\hat{\eta} - \eta_0)$ is asymptotically normal. Let $\tilde{w}(s,t)$ be the limit of $w(s,t)$. Assume that $|w(s,t) - \tilde{w}(s,t)| = O_p(n^{-1/2})$.

Theorem 1. Under condition 1, $\hat{\alpha}_j \to \alpha_0$ $(j = 1, 2)$.

Theorem 2. Under conditions 1 and 2, $n^{1/2}(\hat{\alpha}_j - \alpha_0)$ $(j = 1, 2)$ converge to mean 0 normal random variables.

The proofs of theorem 1 and theorem 2 are given in Appendices B and C respectively. On the basis of the expression

$$n^{1/2}(\hat{\alpha}_j - \alpha_0) = n^{-1/2} \sum_{i=1}^{n} \xi_i(\alpha_0, \eta_0) + o_p(1),$$

where $\xi_i(\alpha_0, \eta_0)$ $(i = 1, \ldots, n)$ are IID mean 0 random variables, we can estimate $\text{var}(\hat{\alpha}_j)$ by $\Sigma_{i=1}^{n} \xi_i^2(\hat{\alpha}, \hat{\eta})/n$. However, an analytical derivation of $\xi_i(\alpha_0, \eta_0)$ is tedious and the expression can be very complicated. Therefore we suggest using resampling methods to obtain a variance estimator. Here we recommend the jackknife approach since in our simulation analysis it produced more reliable confidence intervals than the bootstrap method. Specifically the jackknife estimator of $\text{var}(\hat{\alpha}_1)$ is given by

$$\frac{n-1}{n} \sum_{i=1}^{n} (\hat{\alpha}_1^{(i)} - \hat{\alpha}_1^{(\cdot)})^2$$

where $\hat{\alpha}_1^{(i)}$ is the delete one estimator of $\hat{\alpha}_1$ by leaving the ith observation out and

$$\hat{\alpha}_1^{(\cdot)} = \sum_{i=1}^{n} \hat{\alpha}_1^{(i)}/n.$$

3. Inference procedures based on model 2

Under model 2, the joint distribution of (T_1, T_2) has the form (5) in the upper wedge. If the model is an AC model in equation (8) with $\theta(s, t) = \tilde{\theta}_\alpha\{F(s, t)\}$, the methods discussed in Section 2 can be applied directly. For a copula model that is not in the AC family, $\theta_{\alpha,\eta}(s, t) = A_\alpha\{F_1(s; \alpha), F_2(t)\}$, where $F_1(s; \alpha) = g_\alpha\{F_b(s), F_2(s)\}$, $g_\alpha(\cdot, \cdot)$ satisfies the equation

$$w = C_\alpha\{g_\alpha(w, v), v\} \qquad (0 \leqslant w \leqslant v \leqslant 1) \tag{18}$$

and

$$A_\alpha(u, v) = \frac{C_\alpha(u, v)\, C_\alpha^{11}(u, v)}{C_\alpha^{10}(u, v)\, C_\alpha^{01}(u, v)}, \tag{19}$$

with

$$C_\alpha^{11}(u, v) = \partial^2 C_\alpha(u, v)/\partial u\, \partial v,$$

$$C_\alpha^{10}(u, v) = -\partial C_\alpha(u, v)/\partial u,$$

$$C_\alpha^{01}(u, v) = -\partial C_\alpha(u, v)/\partial v.$$

In this case, $\eta = \{F_b(s), F_2(t)\}$ $(s \leqslant t)$ which can be estimated by Kaplan–Meier estimators. Although we can estimate $\theta_{\alpha,\eta}(s, t)$ by $A_\alpha[g_\alpha\{\hat{F}_b(s), \hat{F}_2(s)\}, \hat{F}_2(t)]$, the resulting estimating function may be too complicated such that conditions 1 and 2 are not satisfied. See Table 2 for selected examples. Therefore we explore other alternatives.

We first examine a 'pseudolikelihood' approach for copula models defined on $[0, \infty)^2$. This approach has been used by Genest *et al.* (1995), Shih and Louis (1995) and Wang and Ding (1999) on the basis of complete data, right-censored data and bivariate current status data respectively.

Table 2. Useful expressions for selected copula models†

	Clayton's model	*Frank's model*	*Positive stable frailty model*
$C_\alpha(u,v)$	$(u^{1-\alpha}+v^{1-\alpha}-1)^{1/(1-\alpha)}$	$\log_\alpha\{1+(\alpha^u-1)(\alpha^v-1)/(\alpha-1)\}$	$\exp(-[\{-\log(u)\}^{1/\alpha}+\{-\log(v)\}^{1/\alpha}]^\alpha)$
$\phi_\alpha(v)$	$(v^{1-\alpha}-1)/(\alpha-1)$	$\log\{(1-\alpha)/(1-\alpha^v)\}$	$\{-\log(v)\}^{1/\alpha}$
$\tilde\theta_\alpha(v)$	α	$-v\log\{\alpha/(1-\alpha^v)\}$	$\tilde\theta_\alpha=(1-\alpha)/[\alpha\{-\log(v)\}]+1$
$g_\alpha(w,v)$	$(w^{1-\alpha}-v^{1-\alpha}+1)^{1/(1-\alpha)}$	$\log_\alpha\{1+(\alpha-1)(\alpha^w-1)/(\alpha^v-1)\}$	$\exp(-[\{-\log(w)\}^{1/\alpha}-\{-\log(v)\}^{1/\alpha}]^\alpha)$
$A_\alpha(u,v)$	α	$\log\{1+Q_\alpha(u,v)\}/Q_\alpha(u,v)$	$1-(\alpha-1)/\alpha[\{-\log(u)\}^{1/\alpha}+\{-\log(v)\}^{1/\alpha}]^\alpha$
Range of α	$\alpha>1$	$\alpha>0$	$0<\alpha<1$

†$C_\alpha(u,v)$, $\phi_\alpha(v)$, $\tilde\theta_\alpha(v)$, $g_\alpha(w,v)$ and $A_\alpha(u,v)$ are defined in equations (5), (8), (9), (18) and (19) respectively.

The main idea is to use nonparametric estimators of $U=F_1(T_1)$ and $V=F_2(T_2)$ as 'pseudo-observations' in the likelihood equation derived on the basis of $\mathrm{pr}(U\geqslant u, V\geqslant v)=C_\alpha(u,v)$. Given data (1), the log-likelihood function can be written as

$$\begin{aligned}l(\alpha,F_1,F_2)=&\sum_{i=1}^n\delta_{1i}\delta_{2i}(\log[C_\alpha^{11}\{F_1(x_i),F_2(y_i)\}]+\log\{f_1(x_i)\}+\log\{f_2(y_i)\})\\&+\sum_{i=1}^n\delta_{1i}(1-\delta_{2i})(\log[C_\alpha^{10}\{F_1(x_i),F_2(y_i)\}]+\log\{f_1(x_i)\})\\&+\sum_{i=1}^n(1-\delta_{1i})\delta_{2i}(\log[C_\alpha^{01}\{F_1(x_i),F_2(y_i)\}]+\log\{f_2(x_i)\})\\&+\sum_{i=1}^n(1-\delta_{1i})(1-\delta_{2i})\log[C_\alpha\{F_1(x_i),F_2(y_i)\}],\end{aligned}$$

where $f_j(\cdot)$ is the density function of T_j $(j=1,2)$. Recall that nonparametric estimators of $F_1(\cdot)$ and $f_1(\cdot)$ do not exist. One possibility is to apply the idea of profile likelihood such that $F_1(s)$ and $f_1(s)$ are replaced by their expressions under the model assumption. Specifically $F_1(s;\alpha,\eta)=g_\alpha\{F_b(s),F_2(s)\}$ and $f_1(s;\alpha,\eta)=-\partial F_1(s)/\partial s$, where $\eta=(F_b(s),F_2(s))$. However, the resulting equation becomes a very complicated function of α and the estimation of $f_1(s;\alpha,\eta)$ involves difficult smoothing problems.

Let us consider applying the Doob–Meyer decomposition under model 2. Define the filtration

$$\begin{aligned}\mathcal{F}_t=\sigma\{&I(X_i\leqslant t,\delta_{1i}=1,\delta_{2i}=0),I(X_i\leqslant t,\delta_{1i}=0,\delta_{2i}=1),I(X_i\leqslant t,\delta_{1i}=1,\delta_{2i}=1),\\&I(Y_i\leqslant t,\delta_{1i}=1,\delta_{2i}=1),\ i=1,\ldots,n\},\end{aligned}$$

which accumulates the information of the end points up to time t. Let $\mathcal{F}=\{\mathcal{F}_t:t\in(0,\infty)\}$. We can derive the compensators

$$E\{\mathrm{d}I(X_i\leqslant s,\delta_{1i}=1)|\mathcal{F}_{s-}\}=I(X_i\geqslant s)\,\Lambda_{10}(\mathrm{d}s;\alpha,\eta)$$

and

$$E\{I(X_i\leqslant t,\delta_{1i}=0,\delta_{2i}=1)|\mathcal{F}_{s-}\}=I(X_i\geqslant s)\,\Lambda_{01}(\mathrm{d}s;\alpha,\eta),$$

where

$$\Lambda_{10}(\mathrm{d}s;\alpha,\eta)=\mathrm{pr}(T_1\in[s,s+\mathrm{d}s)|T_1\geqslant s,T_2\geqslant s)$$
$$=\frac{C_\alpha^{10}\{F_1(s;\alpha,\eta),F_2(s)\}\{-F_1(\mathrm{d}s;\alpha,\eta)\}}{F_b(s)},$$

$$\Lambda_{01}(\mathrm{d}s;\alpha,\eta)=\mathrm{pr}(T_2\in[s,s+\mathrm{d}s)|T_1\geqslant s,T_2\geqslant s)$$
$$=\frac{C_\alpha^{01}\{F_1(s;\alpha,\eta),F_2(s)\}\{-F_2(\mathrm{d}s)\}}{F_b(s)}.$$

The copula assumption without specifying the marginal distributions does not provide enough information to derive the compensator of $I(Y_i\leqslant t,\delta_{1i}=1,\delta_{2i}=1)$ or that of $I(X_i\leqslant t,\delta_{1i}=1,\delta_{2i}=1)$. Two estimating functions of α can be constructed:

$$S_{10}(\alpha,\hat{\eta})=n^{-1}\int w_{10}(s)\{\mathrm{d}\tilde{N}_{10}(s)-\tilde{R}(s)\,\Lambda_{10}(\mathrm{d}s;\alpha,\hat{\eta})\},\qquad(20)$$

$$S_{01}(\alpha,\hat{\eta})=n^{-1}\int w_{01}(s)\{\mathrm{d}\tilde{N}_{01}(s)-\tilde{R}(s)\,\Lambda_{01}(\mathrm{d}s;\alpha,\hat{\eta})\},\qquad(21)$$

where $\hat{\eta}=\{\hat{F}_b(s),\hat{F}_2(t)\}$,

$$\tilde{N}_{10}(t)=\sum_{i=1}^{n}I(X_i\leqslant t,\delta_{1i}=1,\delta_{2i}=0),$$

$$\tilde{N}_{01}(t)=\sum_{i=1}^{n}I(X_i\leqslant t,\delta_{1i}=0,\delta_{2i}=1),$$

$$\tilde{R}(t)=\sum_{i=1}^{n}I(X_i\geqslant t)$$

and $w_{10}(s)$ and $w_{01}(s)$ are weight functions which are predictable with respect to $\mathcal{F}_s$. The solution to $S_*(\alpha)=0$, denoted by $\hat{\alpha}_*$, can be used to estimate α for $*=(10)$ or $*=(01)$. Equations (20) and (21) can be re-expressed as

$$S_*(\alpha,\hat{\eta})=\int_0^t\tilde{w}_*(s)\{\hat{\Lambda}_*^{\mathrm{np}}(\mathrm{d}s)-\hat{\Lambda}_*(\mathrm{d}s;\alpha,\hat{\eta})\}\qquad(*=10,01),$$

where $\tilde{w}_*(s)=\tilde{R}(s)\,W_*(s)/n$ and $\hat{\Lambda}_*^{\mathrm{np}}(t)=\int_0^t\tilde{N}_*(\mathrm{d}s)/\tilde{R}(s)$ for $*=(10)$ or $*=(01)$. Therefore an explanation of $\hat{\alpha}_*$ is that it minimizes the difference between the nonparametric estimator $\hat{\Lambda}_*^{\mathrm{np}}(t)$ and its model-restricted estimator $\hat{\Lambda}_*(t;\alpha,\hat{\eta})$ after appropriate weight adjustment. It should be mentioned that $S_{10}(\alpha,\hat{\eta})$ involves $\hat{F}_1(\mathrm{d}s,\alpha,\hat{\eta})$, which may produce negative mass when $\hat{F}_1(\mathrm{d}s,\alpha,\hat{\eta})$ is not decreasing. Therefore we recommend $S_{01}(\alpha,\hat{\eta})$ for copula models that are not in the AC family.

The asymptotic properties of $\hat{\alpha}_{01}$ rely on the following regularity conditions. To simplify the notation, let $\psi_\alpha(w,v)=C_\alpha^{01}\{g_\alpha(w,v),v\}/w$ and $\psi'_\alpha(w,v)=\partial\psi_\alpha(w,v)/\partial\alpha$.

(a) *Condition 3*: $\psi_\alpha(w,v)$ is bounded, twice differentiable with respect to both w and v with bounded derivatives.

(b) *Condition 4*: $w_{10}(t)$ is predictable with respect to $\mathcal{F}_t$ and is of bounded variation. $n^{1/2n}\{w_{10}(t)-\bar{w}_{10}(t)\}$ converges to a mean 0 Gaussian process, where $\bar{w}_{10}(t)$ is the limit of $w_{10}(t)$ as $n\to\infty$. $\psi_\alpha(w,v)$ is a monotone function of α and is twice differentiable with respect to α with bounded derivatives. $\psi'_\alpha(w,v)$ is twice differentiable with respect to w and v with bounded derivatives.

Uniform and strong consistency of $\Lambda_{01}(t; \alpha_0, \hat{\eta})$ and its weak convergence property can be established on the basis of condition 3. Consistency and asymptotic normality of $\hat{\alpha}_{01}$ can be proved when both condition 3 and condition 4 are valid. The monotonic property of $\psi_\alpha(w, v)$ in α is just a sufficient condition. Consistency holds as long as $S_{01}(\alpha, \hat{\eta}) = 0$ produces a unique root. The proofs use standard techniques and hence are omitted. The asymptotic variance of $\hat{\alpha}_{01}$ can be estimated by using the jackknife method.

4. Numerical analysis

4.1. Simulation studies

In this section, we evaluate the finite sample performance of the estimators proposed. The failure times (T_1, T_2) were generated from the Clayton family (Clayton, 1978) and the Frank family, both of which satisfy model 1 and model 2. We used the algorithm in Prentice and Cai (1992) for the Clayton family and that in Genest (1987) for the Frank family. The independent censoring variable C was generated from a uniform distribution making $\mathrm{pr}(T_1 \leqslant T_2) \approx 0.5$, $\mathrm{pr}(\delta_1 = 0) \approx 0.55$ and $\mathrm{pr}(\delta_2 = 0) \approx 0.1$. Two sample sizes with $n = 150$ and $n = 250$ were chosen. For each estimator, the sample mean and standard deviation of the copula association parameter are reported on the basis of 500 replications. In each case, an estimator of Kendall's τ was derived on the basis of equation (6). Note that τ does not have its original interpretation stated in equation (7) if (T_1, T_2) in the lower wedge follow a different distribution. Nevertheless its value still reveals the strength of association in the upper wedge. We estimated $G(t)$ by the Kaplan–Meier method based on $\{(Y_i, 1 - \delta_{2i})\ (i = 1, \ldots, n)\}$.

Table 3 summarizes the results for the Clayton model with $\theta(s, t) = \alpha$. Three estimators of α with $w(s, t) = w_{01}(t) = 1$ were evaluated. Let $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_{01}$ be the estimators solving equations (16), (17) and (21) respectively. $\hat{\alpha}_1$ is actually the estimator proposed by Day *et al.* (1997).

Table 3. Summary statistics for a simulation based on 500 runs for the Clayton model†

n		*Results for the following values of* τ*:*		
		$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
150	$\hat{\alpha}_{01}$	7.03 (2.79)	11.67 (4.52)	39.91 (12.38)
	$\hat{\alpha}_1$	2.09 (2.60)	1.49 (4.09)	14.58 (8.84)
	$\hat{\alpha}_2$	4.66 (2.63)	4.83 (4.14)	19.72 (9.07)
	$\hat{\tau}_{01}$	10.75 (6.43)	8.46 (5.27)	8.15 (5.17)
	$\hat{\tau}_1$	−0.54 (6.28)	−3.30 (5.12)	1.57 (3.80)
	$\hat{\tau}_2$	5.65 (6.22)	0.80 (5.11)	3.59 (3.86)
250	$\hat{\alpha}_{01}$	5.10 (2.20)	9.98 (3.45)	24.36 (7.25)
	$\hat{\alpha}_1$	1.69 (2.11)	3.07 (3.27)	7.46 (6.25)
	$\hat{\alpha}_2$	3.26 (2.13)	4.95 (3.32)	9.97 (6.32)
	$\hat{\tau}_{01}$	8.45 (5.11)	8.78 (4.05)	7.40 (3.10)
	$\hat{\tau}_1$	0.44 (5.04)	0.57 (4.04)	0.79 (2.76)
	$\hat{\tau}_2$	4.21 (5.04)	2.80 (4.07)	1.86 (2.76)

†In each cell of the first three rows, the first number is the bias ($\times 10^2$) and the number in parentheses is the standard deviation ($\times$ 10) of the estimator of α. In each cell of the last three rows, the first number is the bias ($\times 10^3$) and the number in parentheses ($\times 10^2$) is the standard deviation of the estimator of τ.

Kendall's τ is estimated by using $\tau = (\alpha - 1)/(\alpha + 1)$. The resulting estimators of τ are denoted $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}_{01}$ accordingly. Table 2 shows that $\hat{\alpha}_1$ has the best performance and $\hat{\alpha}_{01}$ has the worst performance in terms of the bias and standard deviation in both sample sizes. The medians of the three estimators also reflected the same pattern as the means so this information is not presented. The result that $\hat{\alpha}_1$, on average, is more accurate than $\hat{\alpha}_2$ is supported by the fact that $L_1(\alpha_0)$ is unbiased whereas $L_2(\alpha_0)$ is not. An explanation why $\hat{\alpha}_{01}$ is inferior in all the cases may be related to the fact that it utilizes only partial information (i.e. $N_{01}(t)$) in the whole data set to estimate the association parameter. Given that $\hat{\tau} = h(\hat{\alpha})$, the delta method implies that $\sigma^2_{\hat{\tau}} \approx h'(\hat{\alpha})^2 \sigma^2_{\hat{\alpha}}$. For Clayton's model with $\alpha > 1$, $h'(\alpha) = 2/(\alpha + 1)^2$ is a decreasing function of α. This explains why, as τ and α increase, $\sigma^2_{\hat{\tau}}$ decreases whereas $\sigma^2_{\hat{\alpha}}$ increases.

Table 4 summarizes the results for the Frank model. We reparameterized the model and directly estimated $\gamma = -\log(\alpha)$. Kendall's τ was estimated using $\tau = 1 + 4\{D_1(\gamma) - 1\}/\gamma$, where

$$D_1(\gamma) = \int_0^\gamma t/\gamma\{\exp(t) - 1\}\,\mathrm{d}t.$$

The previous notation was modified to denote the estimators of γ. In computing $\hat{\gamma}_1$ and $\hat{\gamma}_2$, we used

$$\hat{\theta}_\gamma(s, t) = \gamma\,\hat{F}(s, t)/[1 - \exp\{-\gamma\,\hat{F}(s, t)\}]$$

where

$$\hat{F}(s, t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t)/n\,\hat{G}(t)$$

and $\hat{G}(t)$ is the Kaplan–Meier estimator of $G(t)$ based on $\{(Y_i, 1 - \delta_{2i})\,(i = 1, \ldots, n)\}$. In most cases, $\hat{\gamma}_1$ on average is still the best estimator among its competitors. Somewhat surprisingly,

Table 4. Summary statistics for a simulation based on 500 runs for the Frank model†

n		*Results for the following values of* τ*:*		
		$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
150	$\hat{\gamma}_{01}$	1.44 (9.47)	3.92 (11.78)	13.31 (21.11)
	$\hat{\gamma}_1$	−0.66 (9.53)	−0.77 (10.86)	−1.60 (16.75)
	$\hat{\gamma}_2$	1.91 (9.59)	1.86 (11.13)	1.70 (17.14)
	$\hat{\tau}_{01}$	7.15 (8.19)	14.40 (6.03)	21.06 (3.88)
	$\hat{\tau}_1$	−11.19 (8.50)	−9.96 (6.09)	−7.88 (3.80)
	$\hat{\tau}_2$	11.1 (8.27)	4.27 (5.89)	−0.65 (3.70)
250	$\hat{\gamma}_{01}$	1.40 (7.22)	3.18 (9.17)	8.83 (14.85)
	$\hat{\gamma}_1$	0.20 (7.15)	0.47 (8.70)	0.64 (13.03)
	$\hat{\gamma}_2$	1.90 (7.16)	2.24 (8.78)	2.96 (13.16)
	$\hat{\tau}_{01}$	8.98 (6.22)	13.01 (4.79)	15.12 (2.92)
	$\hat{\tau}_1$	−1.41 (6.30)	−1.11 (4.73)	−1.13 (2.84)
	$\hat{\tau}_2$	13.30 (6.15)	8.35 (4.62)	3.77 (2.79)

†In each cell of the first three rows, the first number is the bias (×10) and the number in parentheses is the standard deviation (×10) of the estimator of γ. In each cell of the last three rows, the first number is the bias (× 10^3) and the number in parentheses (×10^2) is the standard deviation of the estimator of τ.

$\hat{\gamma}_2$ still has much larger bias and is more variable than $\hat{\gamma}_1$ although $L_1(\alpha_0, \hat{\gamma})$ is no longer an unbiased estimating function. However, since τ is not a linear function of γ, we find that, when $n = 150$, $\hat{\tau}_2$ outperforms $\hat{\tau}_1$.

For variance estimation, we selected some cases to compare the bootstrap method and the jackknife method. The jackknife method produced quite accurate coverage probabilities but the bootstrap estimators in all cases substantially overestimated the true variance. The robustness of the proposed estimators under model misspecification was also examined. In the case when the true model was Clayton's ($\tau = 0.5$ and $n = 150$) but was fitted by Frank's, the mean (and standard deviation in parentheses) of $\hat{\tau}_{01}$, $\hat{\tau}_1$ and $\hat{\tau}_2$ are 0.517 (0.091), 0.471 (0.095) and 0.489 (0.089) respectively. In the case when the true model was Frank's ($\tau = 0.3$ and $n = 150$) but was fitted by Clayton's, the mean (and standard deviation) of $\hat{\tau}_{01}$, $\hat{\tau}_1$ and $\hat{\tau}_2$ are 0.375 (0.092), 0.213 (0.065) and 0.216 (0.061) respectively. These results imply that model misspecification inflates the bias and variance of the estimators proposed. The effect of using different weight functions was also evaluated. We chose three weight functions:

$$w_1(s, t) = 1,$$

$$w_2(s, t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t)/n,$$

$$w_3(s, t) = n \Big/ \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t).$$

$w_3(s, t)$ puts more weight in the tail region with less data and $w_2(s, t)$ does the opposite. In all the cases $w_2(s, t)$ performed poorly. For Clayton's model, the unweighted estimator tends to perform better than the weighted versions. For Frank's model, using $w_2(s, t)$ sometimes produces a better estimator but the improvement is slight. Since using a poor weight function may have a very negative effect, we suggest using $w_1(s, t)$ for conservativeness. Note that Greenwood and Wefelmeyer (1991) derived asymptotic optimality criteria for martingale estimating equations. However, when nuisance parameters are involved, the problem of finding the optimal weight becomes much more difficult.

4.2. Data analysis

Two data sets were analysed by using the methods proposed. The first data are available in Klein and Moeschberger (1997), page 464. Among 137 patients receiving bone marrow transplants, 81 died after relapse of leukaemia, only two died without relapse and the remaining 54 patients were doubly censored. Define T_1 as the time from transplantation to relapse of leukaemia and T_2 as the time from transplantation to death. Assuming model 1 with $\theta(s, t) = \alpha$, the proposed unweighted estimators are $\hat{\alpha}_{01} = 10.95$ (2.90), $\hat{\alpha}_1 = 8.78$ (2.27) and $\hat{\alpha}_2 = 8.80$ (2.27), where each number in parentheses is the estimated standard deviation of the corresponding estimator by using the jackknife method. Using the relationship $\alpha = (1+\tau)/(1-\tau)$, the estimators of τ are $\hat{\tau}_{01} = 0.833$ (0.041), $\hat{\tau}_1 = 0.795$ (0.044) and $\hat{\tau}_2 = 0.796$ (0.045). To assess the robustness of the estimators, we also analysed the data by assuming Frank's model with $\theta(s, t) = \gamma F(s, t)/[1 - \exp\{-\gamma F(s, t)\}]$, where $\gamma = -\log(\alpha)$. Using $\tau = 1 + 4\{D_1(\gamma) - 1\}/\gamma$ where $D_1(\gamma) = \int_0^\gamma t/\gamma\{\exp(t) - 1\}\,dt$, the estimated values for τ are $\hat{\tau}_{01} = 0.806$ (0.047), $\hat{\tau}_1 = 0.747$ (0.047) and $\hat{\tau}_2 = 0.748$ (0.047). Fine *et al.* (2001) analysed the same data set and applied the goodness-of-fit test of Shih (1998) for model checking. Their analysis showed that the Clayton model fits the data. Their unweighted estimator for α is 8.79 (2.15) and the weighted estimator is 8.61 (2.15), which are both very close to $\hat{\alpha}_1$ and $\hat{\alpha}_2$. All the results indicate that there is high positive correlation between T_1 and T_2.

The second data set is the Stanford heart transplantation data (Crowley and Hu, 1977). Among 103 participants, 69 received transplants, 30 died without a transplant and only four observations were double censored. Let T_1 be the time from acceptance to heart transplantation and T_2 be the time from acceptance to death. Assuming model 1 with $\theta(s, t) = \alpha$, the estimators proposed are $\hat{\alpha}_1 = 1.153\,(0.268)$ and $\hat{\alpha}_2 = 1.159\,(0.264)$, which correspond to $\hat{\tau}_1 = 0.071\,(0.118)$ and $\hat{\tau}_2 = 0.074$ (0.115). $\hat{\alpha}_{01}$ is not reported since it involves calculating $g_\alpha(w, v)$ which is not stable when α is close to the boundary value 1. Assuming Frank's model, the estimators of τ are $\hat{\tau}_1 = 0.080$ (0.130) and $\hat{\tau}_2 = 0.085$ (0.130). All the analyses showed that the waiting time T_1 and the survival time T_2 seem to be uncorrelated.

5. Concluding remarks

Many survival data can be represented by 2×2 tables or counting processes, which naturally provide the motivation for using the log-rank statistic or the Doob–Meyer decomposition to construct estimating functions. Therefore the ideas proposed may still be applicable under different circumstances especially when likelihood-based inference methods fail. The inference procedures proposed only require specifying model assumptions for the upper wedge. We considered model 1 and model 2 which describe different features of a bivariate relationship. Practitioners may use their scientific knowledge to choose an appropriate model for the data at hand. For the bone marrow transplantation data, whether there is a hypothetical relapse event after death is quite controversial (Prentice *et al.*, 1978). In such a case, $F_1(\cdot)$ is not even well defined. Model 1 has the advantage that it provides an intuitive interpretation of the dependence relationship without directly dealing with F_1. In contrast, model 2 makes some implicit assumption on $F_1(\cdot)$ (i.e. $F_1(T_1) \sim U(0, 1)$) although its explicit form is not specified. In the AIDS example, the definition of T_1 for $T_1 > T_2$ is not a problem since CD4 cell counts still can be measured even after AIDS has occurred. Therefore model 2 will not cause any controversy. For model diagnosis, the method of Shih (1998) for assessing Clayton's model can be applied to semi-competing-risks data. For more general models, Oakes (1989), Genest and Rivest (1993) and Wang and Wells (2000) proposed model selection procedures for AC models defined on $[0, \infty)^2$. Since these methods are not directly applied to data (1), further effort is needed to explore model checking techniques.

An important practical concern is to generalize the methods for accommodating the effect of covariates. Let $Z : p \times 1$ be a vector of covariates. Before imposing a model assumption, it is helpful to examine how the covariates affect (T_1, T_2). If the dependence structure is affected by covariates, we can model the effect via $\theta(s, t)$. For example the Clayton assumption may be extended to $\theta(s, t|Z) = \exp(\beta_0 + \beta' Z) = \alpha(Z)$, where $\beta = (\beta_1, \ldots, \beta_p)'$. If only the marginals are influenced by Z, we may extend model 2 as

$$F(s, t|Z) = C_\alpha\{F_1(s|Z), F_2(t|Z)\},$$

where $F_j(t|Z) = \Pr(T_j \geqslant t|Z)$ $(j = 1, 2)$ may be described by the Cox proportional hazards model. If some covariates Z_1 affect the dependence structure whereas others, denoted Z_2, affect the marginals, an AC model may be described as

$$\theta(s, t|Z_1, Z_2) = \tilde{\theta}_{\alpha(Z_1)}\{F(s, t|Z_2)\}.$$

However, statistical inference under the regression setting is quite challenging since there are many parameters to be estimated jointly but only one estimating function.

Before the name appeared, semi-competing-risks data were analysed extensively in the literature under the context of multistate models or illness–death models. Much previous research focused on T_2 by treating T_1 as auxiliary information. For example Lagakos (1976, 1977) conducted parametric analysis for the joint distribution of (T_1, T_2) for $T_1 \leqslant T_2$. Flandre and O'Quigley (1995) proposed a two-stage approach that uses the progression end point to improve the estimation of survival. Different models describing how T_1 affects T_2 have been proposed on the basis of Markov, semi-Markov or proportional hazards assumptions. A detailed review of these methods can be found in Andersen *et al.* (1991) and Klein and Moeschberger (1997). In recent years, there has been some interest in making inference for the progression time T_1. For example Lin *et al.* (1996) and Chang (2000) considered two-sample comparisons based on T_1. As in this paper, the difficulty in their work arises from the problem of dependent censoring.

Acknowledgements

The work in this paper was funded by National Science Council grant 86-2115-M-001-033-T. The author thanks Mr J. Hsieh for performing the data analysis and the Joint Editor and reviewers for helpful suggestions.

Appendix A: Asymptotic theory for two estimators of $\Lambda_2(t|T_1 > s)$

To simplify the notation,

$$N_{11}(s, t) = \sum_{i=1}^{n} I(X_i \leqslant s, \delta_{1i} = 1, Y_i \leqslant t, \delta_{2i} = 1),$$

$$N_{01}(s, t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \leqslant t, \delta_{2i} = 1),$$

$$N_{10}(s, t) = \sum_{i=1}^{n} I(X_i \leqslant s, \delta_{1i} = 1, Y_i \geqslant t)$$

and

$$\tilde{R}(s, t) = \sum_{i=1}^{n} I(X_i \geqslant s, Y_i \geqslant t).$$

We first derive the properties of $\hat{\Lambda}_2(t|T_1 > s, \alpha_0, \eta_0)$ defined in equation (14). To simplify the notation, let

$$F(s, t) = \text{pr}(T_1 \geqslant s, T_2 \geqslant t),$$

$$F_{11}(s, t) = \text{pr}(F_1 \leqslant s, T_2 \leqslant t),$$

$$F_{10}(s, t) = \text{pr}(F_1 \leqslant s, T_2 \geqslant t),$$

$$F_{01}(s, t) = \text{pr}(F_1 \geqslant s, T_2 \leqslant t),$$

$$H(s, t) = \text{pr}(X \geqslant s, Y \geqslant t),$$

$$H_{11}(s, t) = \text{pr}(X \leqslant s, Y \leqslant t, \delta_1 = 1, \delta_2 = 1),$$

$$H_{10}(s, t) = \text{pr}(X \leqslant s, Y \geqslant t, \delta_1 = 1)$$

and

$$H_{01}(s, t) = \text{pr}(X \geqslant s, Y \leqslant t, \delta_2 = 1).$$

Since C is independent of T_1 and T_2, it follows that

$$\theta_{\alpha_0,\eta_0}(s,t) = \frac{F_{11}(\mathrm{d}s,\mathrm{d}t)/F_{10}(\mathrm{d}s,t)}{F_{01}(s+,\mathrm{d}t)/F(s+,t)} = \frac{H_{11}(\mathrm{d}s,\mathrm{d}t)/H_{10}(\mathrm{d}s,t)}{H_{01}(s+,\mathrm{d}t)/H(s+,t)}, \tag{22}$$

and therefore

$$\Lambda_2(t|T_1 > s) = \int_0^t \frac{H_{01}(s,\mathrm{d}v)}{\theta_{\alpha_0,\eta_0}(s,v)\,H_{10}(\mathrm{d}s,v) + H(s,v) - H_{10}(\mathrm{d}s,v)}. \tag{23}$$

Letting $N_{11}(s,t)/n = \hat{H}_{11}(s,t)$, $N_{10}(s,t)/n = \hat{H}_{10}(s,t)$ and $N_{01}(s,t)/n = \hat{H}_{01}(s,t)$ we can write

$$\hat{\Lambda}_2(t|T_1 > s,\alpha_0,\eta_0) = \int_0^t \frac{\hat{H}_{01}(s,\mathrm{d}v)}{\theta_{\alpha_0,\eta_0}(s,v)\,\hat{H}_{10}(\mathrm{d}s,v) + \hat{H}(s,v) - \hat{H}_{10}(\mathrm{d}s,v)}.$$

To simplify the notation, define $K_n(s,v)$ as

$$\{\theta_{\alpha_0,\eta_0}(s,v)\,H_{10}(\mathrm{d}s,v) + H(s,v) - H_{10}(\mathrm{d}s,v)\}\{\theta_{\alpha_0,\eta_0}(s,v)\,\hat{H}_{10}(\mathrm{d}s,v) + \hat{H}(s,v) - \hat{H}_{10}(\mathrm{d}s,v)\}.$$

Straightforward calculations give

$$\hat{\Lambda}_2(t|T_1 > s,\alpha_0,\eta_0) - \Lambda_2(t|T_1 > s) = a_{1n} + a_{2n} + a_{3n} + a_{4n} + a_{5n} + a_{6n},$$

where

$$a_{1n} = \int_0^t \frac{\theta_{\alpha_0,\eta_0}(s,v)\{H_{10}(\mathrm{d}s,v) - \hat{H}_{10}(\mathrm{d}s,v)\}}{K_n(s,v)}\hat{H}_{01}(s,\mathrm{d}v),$$

$$a_{2n} = \int_0^t \frac{\theta_{\alpha_0,\eta_0}(s,v)\,\hat{H}_{10}(\mathrm{d}s,v)}{K_n(s,v)}\{\hat{H}_{01}(s,\mathrm{d}v) - H_{01}(s,\mathrm{d}v)\},$$

$$a_{3n} = \int_0^t \frac{H(s,v) - \hat{H}(s,v)}{K_n(s,v)}\hat{H}_{01}(s,\mathrm{d}v),$$

$$a_{4n} = \int_0^t \frac{\hat{H}(s,v)}{K_n(s,v)}\{\hat{H}_{01}(s,\mathrm{d}v) - H_{01}(s,\mathrm{d}v)\},$$

$$a_{5n} = -\int_0^t \frac{H_{10}(\mathrm{d}s,v) - \hat{H}_{10}(\mathrm{d}s,v)}{K_n(s,v)}\hat{H}_{01}(s,\mathrm{d}v),$$

$$a_{6n} = -\int_0^t \frac{\hat{H}_{10}(\mathrm{d}s,v)}{K_n(s,v)}\{\hat{H}_{01}(s,\mathrm{d}v) - H_{01}(s,\mathrm{d}v)\}.$$

By the Glivenko–Cantelli theorem, one can show that

$$\sup_{0\leqslant v\leqslant t}|K_n(s,v) - K(s,v)| \to 0 \qquad \text{almost surely,}$$

where

$$K(s,v) = \left\{H(s+,v)\frac{F_{01}(s+,\mathrm{d}v)}{F_{01}(s,\mathrm{d}v)}\right\}^2,$$

which reduces to $H(s+,v)^2$ if $\partial F_{01}(s,v)/\partial v$ is continuous at s. If $H(s+,t) > 0$, it follows that, for all $v \leqslant t < \infty$, there is a constant M such that $1/K_n(s,v) \leqslant M < \infty$ almost surely. Let $\mathcal{T}_s : \sup_t\{t : H(s+,t) > 0\}$. It follows that

$$\sup_{t\in\mathcal{T}_s}|\hat{\Lambda}_2(t|T_1 > s,\alpha_0,\eta_0) - \Lambda_2(t|T_1 > s)| \leqslant \sup_{t\in[0,\mathcal{T}_s]}|a_{3n}| + \sup_{t\in[0,\mathcal{T}_s]}|a_{4n}| + \sup_{t\in[0,\mathcal{T}_s]}|r_n|,$$

where $r_n = a_{1n} + a_{2n} + a_{5n} + a_{6n}$, each of which is of $o_p(n^{-1/2})$. By the Glivenko–Cantelli theorem and applying the techniques in Wang and Wells (1997), pages 876–877, it can be shown that, for any $\varepsilon > 0$, the set

$$\left\{\omega : \sup_{t\in[0,\mathcal{T}_s]} |\hat{\Lambda}_2(t,\omega|T_1 > s, \alpha_0, \eta_0) - \Lambda_2(t|T_1 > s)| > \varepsilon\right\}$$

has measure zero, where ω denotes an element in the corresponding probability space. These results prove that $\hat{\Lambda}_2(t,\omega|T_1 > s, \alpha_0, \eta_0)$ is strongly and uniformly consistent.

Applying integration by parts to a_{4n}, we can write

$$\begin{aligned} n^{1/2}\{\hat{\Lambda}_2(t|T_1 > s, \alpha_0, \eta_0) - \Lambda_2(t|T_1 > s)\} = &-\int_0^t \frac{n^{1/2}\{\hat{H}(s,v) - H(s,v)\}}{K(s,v)} H_{01}(s,\mathrm{d}v) \\ &+ \frac{H(s,t)}{K(s,t)} n^{1/2}\{\hat{H}_{01}(s,t) - H_{01}(s,t)\} \\ &- \int_0^t n^{1/2}\{\hat{H}_{01}(s,v) - H_{01}(s,v)\}\,\mathrm{d}\frac{H(s,v)}{K(s,v)} + o_p(1). \end{aligned}$$

Let $P(s,t)$ and $Q(s,t)$ be the limiting distributions of $n^{1/2}\{\hat{H}(s,t) - H(s,t)\}$ and $n^{1/2}\{\hat{H}_{01}(s,t) - H_{01}(s,t)\}$ respectively, both of which are mean 0 Gaussian processes. The limiting distribution of $n^{1/2}\{\hat{\Lambda}_2(t|T_1 > s, \alpha_0, \eta_0) - \Lambda_2(t|T_1 > s)\}$ becomes

$$-\int_0^t \frac{P(s,v)}{K(s,v)} H_{01}(s,\mathrm{d}v) + \frac{Q(s,t)H(s,t)}{K(s,t)} - \int_0^t Q(s,v)\,\mathrm{d}\frac{H(s,v)}{K(s,v)},$$

which is also a mean 0 Gaussian process. Writing $\tilde{\Lambda}_2(t|T_1 > s)$ in equation (15) as $\int_0^t \hat{H}_{01}(s+,\mathrm{d}v)/\hat{H}(s+,v)$, we can apply similar arguments to show strong and uniform consistency and asymptotic normality of $\tilde{\Lambda}_2(t|T_1 > s)$.

Appendix B: Consistency of $\hat{\alpha}_j$ ($j = 1, 2$)

Assume that the regularity conditions 1 hold. To simplify the notation, define

$$E^*(t; s, \alpha, \eta) = \frac{\theta_{\alpha,\eta}(s,t)\, N_{10}(\mathrm{d}s,t)}{\theta_{\alpha,\eta}(s,t)\, N_{10}(\mathrm{d}s,t) + R(s,t) - N_{10}(\mathrm{d}s,t)}.$$

Note that $\tilde{E}_{11}(\mathrm{d}s,\mathrm{d}t|\alpha,\eta)$ in equation (10) equals $E^*(t;s,\alpha,\eta)\, N_{01}(s,\mathrm{d}t)$. We can write $L_1(\alpha,\eta) = \int_s L^*(\infty; s,\alpha,\eta)$, where

$$L^*(t;s,\alpha,\eta) = n^{-1}\sum_{i=1}^n \int_s^t w(s,u)\, I(X_i \geqslant s)\{I(X_i \leqslant s, \delta_{1i} = 1) - E^*(u;s,\alpha,\eta)\}\,\mathrm{d}I(Y_i \leqslant u, \delta_{2i} = 1).$$

Now we show that $L^*(t;s,\alpha_0,\eta_0)$ is a zero-mean martingale with respect to $\mathcal{F}^*_{s,t}$ in equation (12). Define $M_i(t;s) = I(Y_i \leqslant t, \delta_{2i} = 1) - v_i(t;s)$, where $v_i(t;s)$ is the compensator defined in equation (13). Some algebraic work gives

$$\sum_{i=1}^n I(X_i \geqslant s)\{I(X_i \leqslant s, \delta_{1i} = 1) - E^*(t;s,\alpha_0,\eta_0)\}\,\mathrm{d}v_i(t;s) = 0,$$

and hence we can rewrite $L^*(t;s,\alpha_0,\eta_0)$ as

$$n^{-1}\sum_{i=1}^n \int_s^t w(s,t)\, I(X_i \geqslant s)\{I(X_i \leqslant s, \delta_{1i} = 1) - E^*(u;s,\alpha_0,\eta_0)\}\,\mathrm{d}M_i(u;s).$$

Since the integrand of this expression is $\mathcal{F}^*_{s,t-}$ measurable, $E\{L^*(t;s,\alpha_0,\eta_0)|\mathcal{F}^*_{s,t-}\} = 0$ for each $t > s$ and $E\{L_1(\alpha_0,\eta_0)\} = 0$. This result generalizes lemma 2 in Day *et al.* (1997).

When $\theta_{\alpha,\eta}(s,t)$ is continuous in η and $\hat{\eta} \to^{\mathrm{p}} \eta_0$, it is easy to see that $L_1(\alpha_0,\hat{\eta}) = L_1(\alpha_0,\eta_0) + o_p(1) \to^{\mathrm{p}} 0$. Since $\hat{\Lambda}_2(t|T_1 > s, \alpha_0, \eta_0) = \tilde{\Lambda}_2(t|T_1 > s)$ almost surely, $L_2(\alpha_0,\hat{\eta}) = L_1(\alpha_0,\eta_0) + o_p(1) \to^{\mathrm{p}} 0$. When $\theta_{\alpha,\eta}(s,t)$ is a monotone function of α, it follows that $E^*(t;s,\alpha,\eta)$ and $L_j(\alpha,\hat{\eta})$ ($j = 1, 2$) are both monotone functions of α. In such a case, $L_j(\alpha,\hat{\eta}) = 0$ has the unique solution $\hat{\alpha}_j$ ($j = 1, 2$). When $L_j(\alpha,\eta)$ is twice differentiable with respect to α with bounded derivatives, applying Taylor expansions one can show that

$$\hat{\alpha}_j - \alpha_0 \overset{\text{a}}{=} -\left\{ \frac{\partial L_j(\alpha, \hat{\eta})}{\partial \alpha}\bigg|_{\alpha_0} \right\}^{-1} L_j(\alpha_0, \eta_0) \xrightarrow{\text{p}} 0 \qquad (j = 1, 2).$$

Consistency of $\hat{\alpha}_j$ $(j = 1, 2)$ follows.

Appendix C: Asymptotic normality of $\hat{\alpha}_j$ $(j = 1, 2)$

Assume that the regularity conditions 1 and 2 hold. Here we show asymptotic normality of $n^{1/2}(\hat{\alpha}_2 - \alpha_0)$. Asymptotic normality of $n^{1/2}(\hat{\alpha}_1 - \alpha_0)$ can be proved by using similar arguments. On the basis of equations (22) and (23) in Appendix A, it is easy to see that for any function $w^*(s, t)$

$$\int_s \int_{s \leqslant t} w^*(s, t) \{H_{11}(\mathrm{d}s, \mathrm{d}t) - \theta_{\alpha_0, \eta_0}(s, t)\, H_{10}(\mathrm{d}s, t)\, \Lambda_2(\mathrm{d}t | T_1 > s)\} = 0.$$

Let $\tilde{w}(s, t)$ be the limit of $w(s, t)$ as $n \to w(s, t)$. We can write

$$n^{1/2}(\hat{\alpha}_2 - \alpha_0) \overset{\text{a}}{=} -\left\{ \frac{\partial L_2(\alpha, \eta_0)}{\partial \alpha}\bigg|_{\alpha_0} \right\}^{-1} n^{1/2}\, L_2(\alpha_0, \hat{\eta}),$$

where $n^{1/2}\, L_2(\alpha_0, \hat{\eta})$ can be further expressed as the sum of the terms

$$b_{1n} = \int \int_{s \leqslant t} n^{1/2}\{w(s, t) - \tilde{w}(s, t)\}\{\hat{H}_{11}(\mathrm{d}s, \mathrm{d}t) - \theta_{\alpha_0, \hat{\eta}}(s, t)\, \hat{H}_{10}(\mathrm{d}s, t)\, \tilde{\Lambda}_2(\mathrm{d}t | T_1 > s)\},$$

$$b_{2n} = -\int \int_{s \leqslant t} \tilde{w}(s, t) n^{1/2}\{\theta_{\alpha_0, \hat{\eta}}(s, t) - \theta_{\alpha_0, \eta_0}(s, t)\}\, \hat{H}_{10}(\mathrm{d}s, t)\, \tilde{\Lambda}_2(\mathrm{d}t | T_1 > s),$$

$$b_{3n} = -\int \int_{s \leqslant t} \tilde{w}(s, t)\, \theta_{\alpha_0, \eta_0}(s, t)\, \hat{H}_{10}(\mathrm{d}s, t) n^{1/2}\{\tilde{\Lambda}_2(\mathrm{d}t | T_1 > s) - \Lambda_2(\mathrm{d}t | T_1 > s)\},$$

$$b_{4n} = \int \int_{s \leqslant t} \tilde{w}(s, t) n^{1/2}\{\hat{H}_{11}(\mathrm{d}s, \mathrm{d}t) - H_{11}(\mathrm{d}s, \mathrm{d}t)\} - \int \int_{s \leqslant t} \tilde{w}(s, t)\, \theta_{\alpha_0, \eta_0}(s, t) n^{1/2}\{\hat{H}_{10}(\mathrm{d}s, t) - H_{10}(\mathrm{d}s, t)\}\, \Lambda_2(\mathrm{d}t | T_1 > s).$$

When $|w(s, t) - \tilde{w}(s, t)| = O_p(n^{-1/2})$ for $(s, t) \in \mathcal{P}$, by consistency of the empirical estimators $\hat{\eta}$ and $\tilde{\Lambda}_2(t | T_1 > s)$, it follows that $|b_{1n}| \to^{\text{p}} 0$. By uniform consistency of $\tilde{\Lambda}_2(t | T_1 > s)$ and $\hat{H}_{10}(s, t)$, one can show that

$$b_{2n} \overset{\text{a}}{=} -\int \int_{s \leqslant t} \tilde{w}(s, t) n^{1/2}\, \{\theta_{\alpha_0, \hat{\eta}}(s, t) - \theta_{\alpha_0, \eta_0}(s, t)\}\, H_{10}(\mathrm{d}s, t)\, \Lambda_2(\mathrm{d}t | T_1 > s),$$

which converges in distribution to the mean 0 random variable

$$-n^{1/2}\, (\hat{\eta} - \eta_0) \int \int_{s \leqslant t} \tilde{w}(s, t)\, \frac{\partial \theta_{\alpha_0, \eta}(s, t)}{\partial \eta}\bigg|_{\eta_0} H_{10}(\mathrm{d}s, t)\, \Lambda_2(\mathrm{d}t | T_1 > s).$$

It follows that

$$b_{3n} \overset{\text{a}}{=} -\int \int_{s \leqslant t} w(s, t)\, \theta_{\alpha_0, \eta_0}(s, t)\, H_{10}(\mathrm{d}s, t) n^{1/2}\{\tilde{\Lambda}_2(\mathrm{d}t | T_1 > s) - \Lambda_2(\mathrm{d}t | T_1 > s)\}.$$

Applying integration by parts to make $n^{1/2}\{\tilde{\Lambda}_2(t | T_1 > s) - \Lambda_2(t | T_1 > s)\}$ appear in the integrand, asymptotically b_{3n} can be further expressed as an integral of a mean 0 Gaussian process and hence is asymptotic normal (Shorack and Wellner, 1986). Asymptotic normality of b_{4n} can be shown by using similar arguments.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1991) *Statistical Models based on Counting Processes*. New York: Springer.

Chang, S. H. (2000) A two-sample comparison for multiple ordered event data. *Biometrics*, **56**, 183–189.

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.

Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant survival data. *J. Am. Statist. Ass.*, **72**, 27–36.

Day, R., Bryant, J. and Lefkopolou, M. (1997) Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika*, **84**, 45–56.

Fine, J. P., Jiang, H. and Chappell, R. (2001) On semi-competing risks data. *Biometrika*, **88**, 907–919.

Flandre, P. and O'Quigley, J. (1995) A two-stage procedure for survival studies with surrogate endpoints. *Biometrics*, **51**, 969–976.

Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*. New York: Wiley.

Genest, C. (1987) Frank's family of bivariate distributions. *Biometrika*, **74**, 549–555.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Genest, C. and Rivest, L.-P. (1993) Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Statist. Ass.*, **88**, 1034–1043.

Greenwood, P. E. and Wefelmeyer, W. (1991) On optimal estimating functions for partially specified process models. In *Estimating Functions* (ed. V. P. Godambe). Oxford: Oxford University Press.

Hsu, L. and Prentice, R. L. (1996) A generalization of the Mantel-Haenszel test to bivariate failure time data. *Biometrika*, **83**, 905–911.

Klein, J. P. and Moeschberger, M. L. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Lagakos, S. W. (1976) A stochastic model for censored-survival data in the presence of an auxiliary variable. *Biometrics*, **32**, 551–559.

Lagakos, S. W. (1977) Using auxiliary variables for improved estimates of survival time. *Biometrics*, **33**, 399–404.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996) Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*, **83**, 381–393.

Oakes, D. (1989) Bivariate survival models induced by frailties. *J. Am. Statist. Ass.*, **84**, 487–493.

Prentice, R. L. and Cai, J. (1992) Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495–512.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Shih, J. H. (1998) A goodness-of-fit for association in a bivariate survival model. *Biometrika*, **85**, 189–200.

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

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

Wang, W. and Ding, A. A. (1999) On assessing the association for bivariate current status data. *Biometrika*, **87**, 879–893.

Wang, W. and Wells, M. T. (1997) Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, **4**, 863–880.

Wang, W. and Wells, M. T. (2000) Model selection and semi-parametric inference for bivariate censored data. *J. Am. Statist. Ass.*, **95**, 62–72; rejoinder, 75–76.