J. R. Statist. Soc. B (2008)
70, *Part* 1, *pp.* 3–20

Regression analysis based on semicompeting risks data

Jin-Jian Hsieh and Weijing Wang

National Chiao-Tung University, Hsin-Chu, Taiwan

and A. Adam Ding

Northeastern University, Boston, USA

[Received July 2005. Final revision July 2007]

Summary. Semicompeting risks data are commonly seen in biomedical applications in which a terminal event censors a non-terminal event. Possible dependent censoring complicates statistical analysis. We consider regression analysis based on a non-terminal event, say disease progression, which is subject to censoring by death. The methodology proposed is developed for discrete covariates under two types of assumption. First, separate copula models are assumed for each covariate group and then a flexible regression model is imposed on the progression time which is of major interest. Model checking procedures are also proposed to help to choose a best-fitted model. Under a two-sample setting, Lin and co-workers proposed a competing method which requires an additional marginal assumption on the terminal event and implicitly assumes that the dependence structures in the two groups are the same. Using simulations, we compare the two approaches on the basis of their finite sample performances and robustness properties under model misspecification. The method proposed is applied to a bone marrow transplant data set.

Keywords: Copula model; Dependent censoring; Model selection; Multiple events data; Transformation model

1. Introduction

Many medical studies involve analysis of multiple end points. Such events may be classified into two types, namely terminal and non-terminal. Death is an example of terminal events in the sense that its occurrence precludes the development of others. Examples of non-terminal events, which are subject to censoring by a terminal event, include disease progression or recurrence. If the relationship between the two events is completely unspecified, the marginal distribution of the time to a non-terminal event is not identifiable owing to possible dependent censoring.

Let X be the time to the non-terminal event of major interest, which is usually a status of disease progression, and let Y be the time to death and C be the time to the external censoring event. Observed variables consist of $\tilde{X} = X \wedge Y \wedge C$, $\tilde{Y} = Y \wedge C$, $\delta_x = I(X \leqslant Y \wedge C)$ and $\delta_y = I(Y \leqslant C)$. Such a data structure was called semicompeting risks data by Fine *et al.* (2001). There has been increasing research attention in developing statistical methods for analysing semicompeting risks data. For example, investigation of the degree of association between the two events has been pursued by Day *et al.* (1997) and Fine *et al.* (2001) in which the Clayton model is assumed and Wang (2003) for a class of copula models.

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

 1369–7412/08/70003

In this paper, we consider regression analysis based on progression time X. Because of dependent censoring, the marginal distribution of X is not identifiable non-parametrically. Under a two-sample setting, Lin *et al.* (1996) and Chang (2000) modelled the marginal effects on both X and Y but did not specify their joint distribution. Specifically Lin *et al.* (1996) considered a bivariate location–shift model and Chang (2000) assumed a bivariate accelerated failure time model. This research direction has been further extended to general regression settings in which the non-terminal event is generalized to be recurrent events (Ghosh and Lin, 2003; Lin and Ying, 2003) whereas death still serves as a terminal event. The technique of artificial censoring is used in these references to handle the problem of dependent censoring. Despite being theoretically appealing, the efficiency of the resulting estimator is affected by the degree of artificial censoring. Furthermore, these methods implicitly assume that the dependence structures for the two groups, or for all the levels of covariates, are the same. In other words, they do not account for the situation that covariates may affect the dependence structure.

Here we adopt a different approach to assessing the covariate effect on progression time under dependent censoring. Without making any assumptions on the marginal distribution of Y, we assume that

$$h(X) = -Z'\theta + \varepsilon, \tag{1}$$

where Z is the $p \times 1$ discrete covariate vector, θ is the $p \times 1$ parameter vector, $h(t)$ is a monotonic increasing function and ε is the error term. The parameter θ which measures the covariate effect on X is of major interest. Model (1) can be classified into two classes. One class assumes that $h(t)$ is a known monotone function but leaves the distribution of ε to be unknown. For example, when $h(t) = t$, the model becomes a location–shift model; when $h(t) = \log(t)$, the model follows an accelerated failure time model. The other class assumes that $h(t)$ is unknown but the distribution of ε is completely specified. Examples of the second class include the Cox proportional hazard model with ε being the Gumbel extreme value distribution and the proportional odds model with ε being the standard logistic distribution.

To handle the problem of non-identifiability, we assume that (X, Y) jointly follow an Archimedean copula (AC) model in the upper wedge $\mathcal{P} = \{(x, y) : 0 < x \leqslant y < \infty\}$ such that

$$\Pr(X \geqslant x, Y \geqslant y) = \phi_\alpha^{-1}[\phi_\alpha\{\Pr(X \geqslant x)\} + \phi_\alpha\{\Pr(Y \geqslant y)\}],$$

where $\phi : [0, 1] \rightarrow [0, \infty]$ has two continuous derivatives on $(0, 1)$ and satisfies $\phi(1) = 0$, $\partial\phi(t)/\partial t < 0$ and $\partial^2\phi(t)/\partial t^2 > 0$ for all $0 < t < 1$. Examples of AC models include the Clayton (1978) model $\phi_\alpha(v) = (v^{-\alpha} - 1)/\alpha$ $(\alpha > 0)$, the Frank model (1979) $\phi_\alpha(v) = \log\{(1-\alpha)/(1-\alpha^v)\}$ $(\alpha > 0)$, the Gumbel (1960) model $\phi_\alpha(v) = \{-\log(v)\}^{\alpha+1}$ $(\alpha > 0)$ and the log-copula model $\phi_\alpha(v) = \{1 - \log(v)/\alpha\gamma\}^{\alpha+1} - 1$ $(\alpha, \gamma > 0)$. In the presence of discrete covariates, we assume separate AC models for each covariate group to account for the possibility that the dependence structures for different groups are different. To simplify the notation, let $F_z(x, y) = \Pr(X \geqslant x, Y \geqslant y | Z = z)$, $F_{x,z}(x) = \Pr(X \geqslant x | Z = z)$ and $F_{y,z}(y) = \Pr(Y \geqslant y | Z = z)$. We assume that

$$F_z(x, y) = \phi_{z,\alpha_z}^{-1}[\phi_{z,\alpha_z}\{F_{x,z}(x)\} + \phi_{z,\alpha_z}\{F_{y,z}(y)\}]. \tag{2}$$

Note that, for different groups, we allow not only a different association parameter α_z but also different forms of $\phi_{z,\alpha_z}(\cdot)$.

The inference method proposed for estimating θ under models (1) and (2) is discussed in Section 2. In Section 3, we propose model checking procedures to verify the copula assumption in model (2) and to select an appropriate regression model in equation (1). Simulation results and data analysis are presented in Section 4. Section 5 contains some concluding remarks.

2. A two-stage inference procedure

Let (X_i, Y_i) $(i=1,\ldots,n)$ be independent realizations of (X, Y) which follow model (2) in the upper wedge. The p-dimensional covariate vector for subject i is denoted as Z_i, which takes discrete values, say $z_1,\ldots,z_K$. Denote $n_k=\Sigma_{i=1}^{n} I(Z_i=z_k)$ as the number of observations for the kth subsample and $n=\Sigma_{k=1}^{K} n_k$. Let C_i $(i=1,\ldots,n)$ be independent and identically distributed realizations of the external censoring variable C which is assumed to be independent of (X, Y). We observe semicompeting risks data $\{(\tilde{X}_i, \delta_{xi}, \tilde{Y}_i, \delta_{yi}, Z_i)\ (i=1,2,\ldots,n)\}$, where $\tilde{X}_i = X_i \wedge Y_i \wedge C_i$, $\tilde{Y}_i = Y_i \wedge C_i$, $\delta_{x,i} = I(X_i \leqslant Y_i \wedge C_i)$ and $\delta_{y,i} = I(Y_i \leqslant C_i)$. The inference procedure proposed contains two steps. The parameters in model (2), namely α_z, $F_{y,z}(y)$, $F_z(x, y)$ and $F_{x,z}(x)$, are estimated in the first stage. In the second stage, the proposed estimating function of θ is constructed on the basis of the estimator of $F_{x,z}(x)$.

2.1. First-stage: estimating nuisance parameters

First we obtain the estimators of $F_z(x, y)$, $F_{y,z}(y)$, $F_{x,z}(x)$, $G(y)=\Pr(C \geqslant y)$ and α_z, which are denoted as $\hat{F}_z(x, y)$, $\hat{F}_{y,z}(y)$, $\hat{F}_{x,z}(x)$, $\hat{G}(y)$ and $\hat{\alpha}_z$ respectively, by applying existing methods in the literature to the subsample with $Z=z$.

For $x \leqslant y$, it follows that $F_z(x, y) = \Pr(\tilde{X} \geqslant x, \tilde{Y} \geqslant y | Z=z)/G(y)$. Hence, using the plug-in approach, $F_z(x, y)$ can be estimated by

$$\hat{F}_z(x, y) = \widehat{\Pr}(\tilde{X} \geqslant x, \tilde{Y} \geqslant y | Z=z)/\hat{G}(y) = \sum_{i=1}^{n} I(\tilde{X}_i \geqslant x, \tilde{Y}_i \geqslant y, Z_i = z) \Big/ n_z\, \hat{G}(y), \tag{3}$$

where

$$\hat{G}(y) = \prod_{u<y} \Big\{1 - \sum_{i=1}^{n} I(\tilde{Y}_i = u, \delta_{yi}=0) \Big/ \sum_{i=1}^{n} I(\tilde{Y}_i \geqslant u)\Big\}.$$

This estimator is based on the assumption that covariates Z do not affect the distribution of censoring variable C. In the situation that the distribution of C depends on discrete covariate Z, $\hat{G}(y)$ can be modified by the corresponding Kaplan–Meier estimator $\hat{G}_z(y)$ which uses only those data points with $Z_i = z$. Similarly the estimator of $F_{y,z}(y)$ is given by

$$\hat{F}_{y,z}(y) = \sum_{i=1}^{n} I(\tilde{Y}_i \geqslant y, Z_i = z) \Big/ n_z\, \hat{G}(y). \tag{4}$$

There are several estimators of α_z based on semicompeting risks data. Assuming the Clayton model in the upper wedge, the estimating function that was proposed by Day *et al.* (1997) was constructed on the basis of 2×2 tables and that proposed by Fine *et al.* (2001) utilized the concordant information for paired observations. Wang (2003) generalized the former approach to general AC models. In the absence of covariates, her estimating function of α can be expressed as

$$L(\alpha, \hat{\eta}) = n^{-1} \int\int_{(x,y)\in\mathcal{P}} w(x, y)\{N_{11}(\mathrm{d}x, \mathrm{d}y) - \tilde{E}_{11}(\mathrm{d}x, \mathrm{d}y; \alpha, \hat{\eta})\}, \tag{5}$$

where $w(x, y)$ is a weight function,

$$\tilde{E}_{11}(\mathrm{d}x, \mathrm{d}y; \alpha, \eta) = \frac{\theta_{\alpha,\eta}(x, y)\, N_{10}(\mathrm{d}x, y)\, N_{01}(x, dy)}{\theta_{\alpha,\eta}(x, y)\, N_{10}(\mathrm{d}x, y) + R(x, y) - N_{10}(\mathrm{d}x, y)},$$

$$N_{11}(\mathrm{d}x, \mathrm{d}y) = \sum_{i=1}^{n} I(\tilde{X}_i = x, \tilde{Y}_i = y, \delta_{xi}=1, \delta_{yi}=1),$$

$$N_{10}(\mathrm{d}x, y) = \sum_{i=1}^{n} I(\tilde{X}_i = x, \delta_{xi}=1, \tilde{Y}_i \geqslant y),$$

$$N_{01}(x, \mathrm{d}y) = \sum_{i=1}^{n} I(\tilde{X}_i \geqslant x, \tilde{Y}_i = y, \delta_{yi} = 1),$$

$$R(x, y) = \sum_{i=1}^{n} I(\tilde{X}_i \geqslant x, \tilde{Y}_i \geqslant y)$$

and $\theta_{\alpha,\eta}(x, y) = \tilde{\theta}_\alpha\{F(x, y)\}$ with

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

and $\eta = F(x, y)$ can be estimated by $\hat{\eta} = \hat{F}(x, y)$ by using formula (3) without further partitioning by Z.

Here we modify Wang's method to estimate α_z by using only data points with $Z_i = z$. Then, on the basis of model (2), we can derive $F_{x,z}(x)$ in terms of $\phi_{z,\alpha_z}(\cdot)$, $F_z(x, y)$ and $F_{y,z}(y)$. Fine *et al.* (2001) suggested to consider the relationship on the diagonal line $y = x$ and, by straightforward calculation, we obtain

$$F_{x,z}(x) = \phi^{-1}_{z,\alpha_z}[\phi_{z,\alpha_z}\{F_z(x, x)\} - \phi_{z,\alpha_z}\{F_{y,z}(x)\}] = H_{z_k}\{F_z(x, x), F_{y,z}(x), \alpha_z\}.$$

The marginal function $F_{x,z}(x)$ can be estimated by

$$\hat{F}_{x,z}(x) = \phi^{-1}_{z,\hat{\alpha}_z}[\phi_{z,\hat{\alpha}_z}\{\hat{F}_z(x, x)\} - \phi_{z,\hat{\alpha}_z}\{\hat{F}_{y,z}(x)\}] = H_{z_k}\{\hat{F}_z(x, x), \hat{F}_{y,z}(x), \hat{\alpha}_z\}. \quad (6)$$

2.2. Second stage: estimating the regression parameter

The proposed estimating equation of θ is motivated by the following two-sample test statistic with $Z = 0, 1$. Specifically to test $F_{x,0}(t) = F_{x,1}(t)$ for every t within the range of the data, we can use

$$U_T = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \int W(x) \left\{\hat{F}_{x,0}(x) - \hat{F}_{x,1}(x)\right\} \mathrm{d}x, \quad (7)$$

where $W(x)$ is a weight function.

Now we modify the test statistic U_T in equation (7) to construct an estimating equation for one-dimensional θ with $Z = 0, 1$. Let θ_0 be the true value of θ. Model (1) induces a functional transformation $\xi_\theta(\cdot)$ such that $\xi_{\theta_0}(F_{x,0}) = F_{x,1}$. When $h(\cdot)$ is known but the distribution of ε is unknown, $\xi_\theta(F)(t) = F[h^{-1}\{h(t) + \theta\}]$; when $h(\cdot)$ is unknown but the distribution of ε is known, $\xi_\theta(F)(t) = F_\varepsilon[F_\varepsilon^{-1}\{F(t)\} + \theta]$, where $F_\varepsilon(t) = \Pr(\varepsilon \geqslant t)$ denotes the survival function of ε. Now we can define a function $g(t, \theta)$ such that

$$g(t, \theta) = \xi_\theta(F_{x,0})(t) - F_{x,1}(t).$$

Then $g(t, \theta_0) = 0$ for all t. Since

$$\sqrt{\left(\frac{n_0 n_1}{n}\right)} \int W(x)\, g(x, \theta_0)\, \mathrm{d}x = 0,$$

we can then estimate θ by solving the corresponding estimating equation

$$U(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \int W(x)\, \hat{g}(x, \theta)\, \mathrm{d}x = 0,$$

where $\hat{g}(t, \theta) = \xi_\theta(\hat{F}_{x,0})(t) - \hat{F}_{x,1}(t)$.

The above idea can be modified to account for the situation that Z contains multiple covariates but all of them have finite discrete values. In such a case, let $\{z_k, k = 1, 2, \ldots, K\}$ denote the set of all possible Z-values. Now z_k, θ and θ_0 are $p\times 1$ vectors. When model (1) is true, it follows that $\xi_{(z_j-z_k)^{\mathrm{T}}\theta_0}(F_{x,z_k})=F_{x,z_j}$. Here and throughout the paper, a^{T} denotes the transpose of a. Define $g_{kj}(t,\theta)=\xi_{z_{kj}^{\mathrm{T}}\theta}(F_{x,z_k})(t)-F_{x,z_j}(t)$ and $\hat{g}_{kj}(t,\theta)=\xi_{z_{kj}^{\mathrm{T}}\theta}(\hat{F}_{x,z_k})(t)-\hat{F}_{x,z_j}(t)$, where $z_{kj}=z_j-z_k$ and $\hat{F}_{x,z_k}$ is the estimator (6) based on the subsample with $Z=z_k$. The estimating function then becomes

$$U(\theta)=\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left(\frac{n_k n_j}{n_k+n_j}\right)}\int_0^{t_{kj}} W_{kj}(t)\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t \tag{8}$$

where $w_0(\cdot)$ and $W_{kj}(\cdot)$ are the weight functions, and t_{kj} is the largest value of $\tilde{X}$ in the pooled subsample with $Z=z_k$ or $Z=z_j$. The proposed estimator of θ is the solution to $U(\theta)=0$, which is denoted as $\hat{\theta}$.

Asymptotic properties of $\hat{\theta}$ which solves $U(\theta)=0$ are given in the following theorem.

Theorem 1. Assume that models (1) and (2) hold. Under the regularity conditions that are stated in Appendix A, $\hat{\theta}$ is a consistent estimator of θ_0 and $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotically normal with mean 0, where θ_0 is the true value.

A sketch of the proof is outlined in Appendix B. More detailed discussions are provided in Hsieh *et al.* (2007). Since it is not easy to estimate the asymptotic variance of $\hat{\theta}$ by an analytic formula, we suggest the use of a bootstrap or a jackknife method to estimate its variance.

In practice, the weight function may also be estimated. Replacing $W_{kj}(t)$ in equation (8) with $\hat{W}_{kj}(t)$, we have the estimating function

$$\hat{U}(\theta)=\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left(\frac{n_k n_j}{n_k+n_j}\right)}\int_0^{t_{kj}} \hat{W}_{kj}(t)\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t.$$

The Gehan-type weights are often used (page 230 of Klein and Moeschberger (2003)); these can be written as

$$\hat{W}_{kj}(x)=\frac{(n_k+n_j)\,\hat{G}_{z_k}(x)\,\hat{G}_{z_j}(x)}{n_k\,\hat{G}_{z_k}(x)+n_j\,\hat{G}_{z_j}(x)},$$

where $\hat{G}_{z_k}(x)$ is the Kaplan–Meier estimator of $G_{z_k}(x)=\Pr(C\geqslant x|Z=z_k)$. Note that $\hat{W}_{kj}(x)$ is an estimator of

$$W_{kj}(x)=\frac{(c_k+c_j)\,G_{z_k}(x)\,G_{z_j}(x)}{c_k\,G_{z_k}(x)+c_j\,G_{z_j}(x)},$$

where c_k and c_j are the constants that are defined in the first regularity condition (a) that is listed in Appendix A. Let $\tilde{\theta}$ solve $\hat{U}(\theta)=0$. Its asymptotic properties are stated in the following theorem. In Appendix C, we present a sketch of the proof and, for the details, refer to Hsieh *et al.* (2007).

Theorem 2. If $\hat{W}_{kj}(t)$ uniformly strongly converges to $W_{kj}(t)$ then, under the conditions for theorem 1, the solution to the estimating equation $\hat{U}(\theta)=0$ is also asymptotically normal, i.e. let $\tilde{\theta}$ denote the solution to $\hat{U}(\theta)=0$; then $n^{1/2}(\tilde{\theta}-\theta_0)$ weakly converges to a mean 0 normal random variable, where θ_0 is the true value.

For computation, we may use the fact that $\hat{F}_{x,0}(t)$ and $\hat{F}_{x,1}(t)$ are piecewise constant functions. Let $t_{(1)}\leqslant\ldots\leqslant t_{(n)}$ be the observed ordered times of $\tilde{X}$ in the pooled sample and set $t_{(0)}=0$. Then $\hat{F}_{x,0}(t)$ and $\hat{F}_{x,1}(t)$ are constants on the time intervals $(t_{(i-1)},t_{(i)}]$. Usually, the estimated weight functions such as the Gehan-type weights can also be taken to be piecewise constant functions between $t_{(i-1)}$ and $t_{(i)}$ which would enable simplification for computation. For example, with piecewise constant weight function $\hat{W}(t)$, the quantity correponding to U_T in equation (7) can be rewritten as

$$\hat{U}_T = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \sum_{i=1}^{n} \hat{W}(t_{(i)})(t_{(i)} - t_{(i-1)})\{\hat{F}_{x,0}(t_{(i)}) - \hat{F}_{x,1}(t_{(i)})\}. \tag{9}$$

For illustration, we now derive the estimating equations under a two-sample setting for selected examples.

2.2.1. *Example 1: Cox proportional hazard model*

When ε has the extreme value distribution, model (1) becomes the Cox proportional hazard model. Then $F_\varepsilon(t) = \exp\{-\exp(t)\}$ and $\xi_\theta(F) = F^{\exp(\theta)}$. When θ equals its true value θ_0, it follows that

$$F_{x,1}(x) = F_{x,0}(x)^{\exp(\theta_0)}.$$

Therefore $g(t,\theta) = F_{x,0}(t)^{\exp(\theta)} - F_{x,1}(t)$, and the estimating equation is

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \int_0^{t_{(n)}} \hat{W}(t)\{\hat{F}_{x,0}(t)^{\exp(\theta)} - \hat{F}_{x,1}(t)\}\,\mathrm{d}t = 0.$$

Under the piecewise constant weight function, the resulting estimating equation becomes

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \sum_{i=1}^{n} \hat{W}(t_{(i)})(t_{(i)} - t_{(i-1)})\{\hat{F}_{x,0}(t_{(i)})^{\exp(\theta)} - \hat{F}_{x,1}(t_{(i)})\} = 0.$$

2.2.2. *Example 2: the proportional odds model*

When ε is the standard logistic distribution, model (1) becomes the proportional odds model, where $F_\varepsilon(t) = 1/\{1+\exp(t)\}$ and $\xi_\theta(F) = F/\{\exp(\theta) - F\ \exp(\theta) + F\}$. When θ equals its true value θ_0, it follows that

$$F_{x,1}(t) = \frac{F_{x,0}(t)}{\exp(\theta_0) - F_{x,0}(t)\exp(\theta_0) + F_{x,0}(t)},$$

and

$$g(t,\theta) = \frac{F_{x,0}(t)}{\exp(\theta) - F_{x,0}(t)\exp(\theta) + F_{x,0}(t)} - F_{x,1}(t).$$

So the estimating equation is

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \int_0^{t_{(n)}} \hat{W}(t) \left\{ \frac{\hat{F}_{x,0}(t)}{\exp(\theta) - \hat{F}_{x,0}(t)\exp(\theta) + \hat{F}_{x,0}(t)} - \hat{F}_{x,1}(t) \right\} \mathrm{d}t = 0.$$

Under the piecewise constant weight function, the resulting estimating equation becomes

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \sum_{i=1}^{n} \hat{W}(t_{(i)})(t_{(i)} - t_{(i-1)}) \left\{ \frac{\hat{F}_{x,0}(t_{(i)})}{\exp(\theta) - \hat{F}_{x,0}(t_{(i)})\exp(\theta) + \hat{F}_{x,0}(t_{(i)})} - \hat{F}_{x,1}(t_{(i)}) \right\} = 0.$$

2.2.3. Example 3: the accelerated failure time model

When $h(t) = \log(t)$, model (1) becomes the accelerated failure time model. Now $\xi_\theta(F)(t) = F\{\exp(\theta)t\}$. When θ equals its true value θ_0, it follows that

$$F_{x,1}(t) = F_{x,0}\{\exp(\theta_0)t\},$$

and

$$g(t,\theta) = F_{x,0}\{\exp(\theta_0)t\} - F_{x,1}(t).$$

So the estimating equation is

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \int \hat{W}(t)[\hat{F}_{x,0}\{\exp(\theta)t\} - \hat{F}_{x,1}(t)]\,\mathrm{d}t = 0,$$

where $\hat{F}_{x,0}\{\exp(\theta)t\} = \widehat{\Pr}\{X \geqslant \exp(\theta)t|Z=0\} = \widehat{\Pr}\{\exp(-\theta)X \geqslant t|Z=0\}$. Note that the discontinuous points of $\hat{F}_{x,0}\{\exp(\theta)t\}$ are different from those of $\hat{F}_{x,0}(t)$. Denote $\hat{F}^*_{x,0}(t)$ as the estimator $\hat{F}_{x,0}(t)$ computed on the basis of the transformed data

$$\{\exp(-\theta)\tilde{X}_i, \exp(-\theta)\tilde{Y}_i, \delta_{xi}, \delta_{yi}\}$$

for i with $Z_i = 0$. Let $\tilde{t}_{(1)} \leqslant \ldots \leqslant \tilde{t}_{(n)}$ be the order times of the pooled sample

$$\{\exp(-\theta)\tilde{X}_i\, I(Z_i = 0) + \tilde{X}_i\, I(Z_i = 1)\ (i = 1, \ldots, n)\}.$$

If the weight function is piecewise constant and takes jumps at $\{\tilde{t}_{(j)}, j = 1, \ldots, n\}$, the resulting estimating function becomes

$$\hat{U}(\theta) = \sqrt{\left(\frac{n_0 n_1}{n}\right)} \sum_{i=1}^{n} \hat{W}(\tilde{t}_{(i)})(\tilde{t}_{(i)} - \tilde{t}_{(i-1)})\{\hat{F}^*_{x,0}(\tilde{t}_{(i)}) - \hat{F}_{x,1}(\tilde{t}_{(i)})\} = 0.$$

3. Model selection

The procedure proposed is developed on the basis of two assumptions: the dependence structure of an AC model characterized by $\phi_{z,\alpha_z}(\cdot)$ in model (2) and the regression model in expression (1). By specifying the dependence relationship between X and Y for each value of Z, we can avoid making unnecessary assumptions about the covariate effect on Y as in Lin *et al.* (1996). Now we discuss how to justify the assumptions imposed.

3.1. Selection of a copula model

We first consider how to check whether a copula model $\phi_{z,\alpha_z}(\cdot)$ fits the data at hand for each covariate group. Without loss of generality and to simplify the presentation, the discussions here are based on a homogeneous sample $\{(\tilde{X}_i, \delta_{xi}, \tilde{Y}_i, \delta_{yi})\ (i = 1, 2, \ldots, n)\}$ such that (X, Y) follows an AC model

$$F(x,y) = C_\alpha\{F_x(x), F_y(y)\} = \phi_\alpha^{-1}[\phi_\alpha\{F_x(x)\} + \phi_\alpha\{F_y(y)\}]. \tag{10}$$

We briefly summarize our ideas. Consider the function $F^{11}(t_1, t_2) = \Pr(X \geqslant t_1, Y \geqslant t_2|\delta_x = 1, \delta_y = 1)$ which is identifiable non-parametrically in the upper wedge $\{(t_1, t_2): 0 < t_1 \leqslant t_2 < \infty\}$. By comparing the non-parametric estimator of $F^{11}(t_1, t_2)$ and its model-based estimator for $F^{11}(t_1, t_2)$ on the basis of some distance measure, we can find the most plausible model which is the model that yields the smallest distance among the candidates. Furthermore a formal goodness-of-fit test can be constructed if the distribution of the distance measure under the null hypothesis can be derived. Since analytic derivations are complicated, we suggest using the bootstrap resampling method to obtain the cut-off value in the test.

The non-parametric estimator, which is denoted as $\hat{F}^{11}(t_1, t_2)$ $(t_1 \leqslant t_2)$, is given by

$$\sum_{i=1}^{n} I(\tilde{X}_i \geqslant t_1, \tilde{Y}_i \geqslant t_2, \delta_{xi}=1, \delta_{yi}=1) \Big/ \sum_{i=1}^{n} I(\delta_{xi}=1, \delta_{yi}=1).$$

Assume that there are K model candidates $C_{\alpha}^{(k)}\{F_x(x), F_y(y)\}$ $(k=1,2,\ldots,K)$, each of which can be characterized by $\phi_{\alpha}^{(k)}$. Note that the definition of α depends on the model chosen. For an AC model that is indexed by $\phi_{\alpha}^{(k)}$, the model-based estimator, which is denoted as $\tilde{F}_k^{11}(t_1, t_2)$, can be computed over the region $\{t_1 \leqslant t_2\}$ as follows:

$$\tilde{F}_k^{11}(t_1, t_2) = \frac{\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} \tilde{F}_k(\mathrm{d}x, \mathrm{d}y)\, \hat{G}(y)}{\int_{y=0}^{\infty} \int_{x=0}^{y} \tilde{F}_k(\mathrm{d}x, \mathrm{d}y)\, \hat{G}(y)},$$

where $\tilde{F}_k(\mathrm{d}x, \mathrm{d}y) = \tilde{F}_k(x, y) - \tilde{F}_k(x+\mathrm{d}x, y) - \tilde{F}_k(x, y+\mathrm{d}y) + \tilde{F}_k(x+\mathrm{d}x, y+\mathrm{d}y)$ and $\tilde{F}_k(x, y) = \phi_{\hat{\alpha}}^{(k)^{-1}}[\phi_{\hat{\alpha}}^{(k)}\{\hat{F}_x(x)\} + \phi_{\hat{\alpha}}^{(k)}\{\hat{F}_y(y)\}]$. To verify whether a copula model $\phi_{\alpha}^{(k)}$ fits the data, we can perform a formal testing procedure as follows. Consider testing $H_0: \phi_{\alpha} = \phi_{\alpha}^{(k)}$ *versus* $H_{\mathrm{a}}: \phi_{\alpha} \neq \phi_{\alpha}^{(k)}$. Define

$$D^k = \sup_{t_1 \leqslant t_2} |\hat{F}^{11}(t_1, t_2) - \tilde{F}_k^{11}(t_1, t_2)|. \tag{11}$$

We can reject H_0 if $D^k > c_k$, where c_k is the critical value satisfying $\Pr(D^k > c_k | H_0) = \gamma$, the prespecified type I error rate.

Because the distribution of D^k is difficult to derive analytically, we suggest using bootstrap resampling methods to obtain the cut-off value, p-value and power. Here we briefly describe the procedure. A bootstrap sample under model $\phi_{\alpha}^{(k)}$ can be generated as follows. Recall that, given the original data, we have obtained $\hat{G}(c)$, $\hat{F}_y(y)$ and $\hat{F}_x(x)$ under the assumption of model $\phi_{\alpha}^{(k)}$. Then generate $(U_i^*, V_i^*) \sim$ copula model k with $U_i^* \sim U(0,1)$ and $V_i^* \sim U(0,1)$. Then set $X_i^* = s$ if $\hat{F}_x(s^+) < 1 - U_i^* \leqslant \hat{F}_x(s)$, $Y_i^* = t$ if $\hat{F}_y(t^+) < 1 - V_i^* \leqslant \hat{F}_y(t)$ and $C_i^* \sim \hat{G}(c)$. Given (X_i^*, Y_i^*, C_i^*) $(i=1,\ldots,n)$, we can construct a bootstrap sample $\{(\tilde{X}_i^*, \delta_{xi}^*, \tilde{Y}_i^*, \delta_{yi}^*)\ (i=1,2,\ldots,n)\}$, where $\tilde{X}_i^* = X_i^* \wedge Y_i^* \wedge C_i^*$, $\tilde{Y}_i^* = Y_i^* \wedge C_i^*$, $\delta_{xi}^* = I(X_i^* \leqslant Y_i^* \wedge C_i^*)$ and $\delta_{yi}^* = I(Y_i^* \leqslant C_i^*)$. With a bootstrapped sample, we can compute the corresponding values of D^k. Repeating the bootstrapping procedure many times, the distribution of D^k can be approximated by the empirical counterparts from the bootstrapped samples.

The above tests will reject the null hypothesis if the data obviously violate the copula model $\phi_{\alpha}^{(k)}$. In practice, we may be more interested in choosing the best-fitted copula model from several candidates that are indexed by $k=1,2,\ldots,K$. For this, we can select the model that yields the smallest D^k. Now we derive theoretical properties of the model selection procedure proposed.

Theorem 3. Assume that (X, Y) follow model (10) and both variables are continuous and the independent censoring variable C has bigger support than the supports of X and Y. Suppose that there are K model candidates in the AC family. Let the kth model $C_{\alpha}^{(k)}(u, v)$ be characterized by $\phi_{\alpha}^{(k)}(t)$, which has regular analytic properties in t and is continuous in α, whose parameter space is a closed set. If $\phi_{\alpha}^{(k)}$ is the true copula model, $D^k \rightarrow^{\mathrm{P}} 0$ as $n \rightarrow \infty$. If $\phi_{\alpha}^{(k)}$ is not the true model, $\Pr\{\liminf_{n\rightarrow\infty}(D^k) > 0\} = 1$. Furthermore let $\hat{k}$ denote the copula model that yields the smallest D^k among all the candidates. Then $\phi_{\alpha}^{(\hat{k})}$ is consistent if the true copula model is included in the list of candidates.

In Appendix D, a sketch of the proof for theorem 3 is given. A more detailed proof can be found in section 3 of Hsieh *et al.* (2007).

3.2. Selection of the covariate model

After specifying the form of model (2), our procedure requires choosing an appropriate regression model in expression (1). If model (1) is correctly specified, $g_{kj}(t,\theta_0)=0$ and it is reasonable to expect that $\hat{g}_{kj}(t,\hat{\theta})$ is closer to zero for the correct model than a wrong model for moderate sample sizes. This fact can be used to check the model assumption (1). Let $D_R=\max_{k,j,t}|\hat{g}_{kj}(t,\hat{\theta})|$. A formal model checking procedure can be formulated as testing the hypothesis H_0: the form of model (1) is correct *versus* H_a: the form of model (1) is not correct. The null hypothesis is rejected if D_R is too big. The cut-off value for the test can be calculated by applying the bootstrapped method which can also be used for model selection. Suppose that there are several choices for model (1), say model $k=1,2,\ldots,K$. To select the best-fitting model, we can simply choose the model with smallest D_R^k, where D_R^k is calculated under model k.

4. Numerical analysis

4.1. Simulation results

We designed several simulation settings to examine the validity and robustness of the methods proposed. Data generation algorithms for the Clayton model and the Frank model have been given in Prentice and Cai (1992) and Genest (1987) respectively. In the following analysis, we set the weight functions as $w_0(z'_{ij}\theta)=1$ and

$$\hat{W}_{ij}(x)=\frac{(n_i+n_j)\,\hat{G}_{z_i}(x)\,\hat{G}_{z_j}(x)}{n_i\,\hat{G}_{z_i}(x)+n_j\,\hat{G}_{z_j}(x)}.$$

For each estimator under evaluation, the average bias and the standard deviation based on 1000 runs are reported. Here we describe only summary information of the numerical settings. For the details, refer to section 4 of Hsieh *et al.* (2007).

Tables 1 and 2 contain the results of the first analysis that compared our proposed estimator $\hat{\theta}$ and $\hat{\theta}_L$, the estimator of Lin *et al.* (1996). We set $(\varepsilon,\xi)|Z$ to follow an AC model with $Z=0,1$. Then, on the basis of (ε,ξ,Z), the value of (X,Y) can be determined from the models $h_1(X)=-\theta_0 Z+\varepsilon$ and $h_2(Y)=-\eta_0 Z+\xi$. Here we set $\theta_0=\eta_0=0.5$ and $n_0=n_1=150$. Note that all the assumptions are satisfied for $\hat{\theta}$. However, in the evaluation of $\hat{\theta}_L$, the covariate model for X is correct but the assumption about common dependence structures for the two groups or the extra assumption on a covariate model for Y may be misspecified.

In the four cases in Table 1, we consider the location–shift model with $h_1(t)=h_2(t)=t$. We shall

Table 1. Finite sample performance of two estimators evaluated under four situations†

Model	$\hat{\theta}$	$\hat{\theta}_L$
Case 1	−0.0026 (0.0934)	−0.0025 (0.0909)
Case 2	−0.0013 (0.1136)	0.0969 (0.0849)
Case 3	0.0022 (0.0950)	−0.0122 (0.0888)
Case 4	0.0008 (0.1100)	0.0982 (0.0840)

†The correlation structures are the same for two covariate groups in the first case and different in the last three cases. The first number is the average bias of the estimator and the number in parentheses is the standard deviation based on 1000 replications.

Table 2. Finite sample performance of two estimators evaluated under four situations with different covariate models for progression time and death time (thus $\hat{\theta}_L$ becomes invalid)†

Model	$\hat{\theta}$	$\hat{\theta}_L$
Case 5	−0.0041 (0.0974)	0.0890 (0.1175)
Case 6	−0.0067 (0.1135)	0.3387 (0.1127)
Case 7	0.0025 (0.1156)	0.0884 (0.1170)
Case 8	0.0125 (0.1152)	0.3793 (0.1081)

†The first number is the average bias of the estimator and the number in parentheses is the standard deviation based on 1000 replications.

use the notation $\{\text{Clayton}(\tau_0), \text{Frank}(\tau_1)\}$ to denote the situation that one group with $Z=0$ follows the Clayton model with $\tau=\tau_0$ and the other with $Z=1$ follows the Frank model with $\tau=\tau_1$. The dependence structures for the four cases are case 1, $\{\text{Clayton}(0.5), \text{Clayton}(0.5)\}$, case 2, $\{\text{Clayton}(0.8), \text{Clayton}(0.1)\}$, case 3, $\{\text{Frank}(0.5), \text{Clayton}(0.5)\}$, and case 4, $\{\text{Frank}(0.8), \text{Clayton}(0.1)\}$. In case 1 where the conditions for both estimators are valid, $\hat{\theta}_L$ slightly outperforms $\hat{\theta}$. However, in the last three cases, $\hat{\theta}_L$ is biased. It seems that the bias of $\hat{\theta}_L$ is affected more by the discrepancy in the level of associations for the two groups than the difference in the dependence structures.

Table 2 contains the results for another four conditions (cases 5–8). We set $h_1(t)=t$ but $h_2(t)=\log(t)$, which is a condition that violates the assumption that was made by Lin *et al.* (1996). The dependence structures in these four cases follow the same pattern as in cases 1–4 in Table 1. We see that $\hat{\theta}$ outperforms $\hat{\theta}_L$ even more since, for the latter, the two types of assumption are both misspecified.

The second analysis checks the validity of the proposed method for selecting an appropriate copula model. Suppose that there are two copula models under consideration, where model $k=1$ is the Clayton model and model $k=2$ is the Frank model. First we set the Clayton model as the true model and $n=150$. The mean and standard deviation (in parentheses) of D^1 and D^2 are 0.0780 (0.0187) and 0.1397 (0.0304) on the basis of 1000 replications. The percentages of successfully selecting the Clayton model are 93.4% on the basis of the order of D^j $(j=1,2)$. Then we set the Frank model as the true model. The mean and standard deviation (in parentheses) of D^1 and D^2 are 0.1398 (0.0330) and 0.0819 (0.0206). The percentages of successfully selecting the Frank model are 92.3% on the basis of the order of D^j $(j=1,2)$. Finally, we examine the proposed testing procedure by using the resampling method. Under the Clayton model, we set up the goodness-of-fit test H_0: the data follow the Clayton model *versus* H_a: the data do not follow the Clayton model. By resampling 1000 times, we obtained $D^1=0.0511$ with p-value 0.909 and the cut-off value $c_1=0.1004$ (at 0.05 significance level). Hence hypothesis H_0 is accepted, which is the correct decision. For the same data set, we ran the analysis again with H_0: the data follow the Frank model *versus* H_a: the data do not follow the Frank model. We obtained that $D^2=0.1247$ with p-value 0.012; the cut-off value $(\gamma=0.05)$ $c_2=0.1058$. Accordingly we reject hypothesis H_0, which is also the correct decision.

Under the Clayton model with $\tau_0=0.5$ and $\tau_1=0.6$, we examine the method proposed that was introduced in Section 3.2 for selecting an appropriate regression model. Table 3 lists the proportions of each model being selected by the method proposed on the basis of 500 simulation runs. The proportion of times that the true model is selected increases as the sample size grows

Table 3. Proportion of the covariate models that were selected by the method proposed on the basis of 500 replications†

True model	*n*	*Proportions (%) for the following chosen models:*			
		Location–shift	*Accelerated failure time*	*Proportional hazards*	*Proportional odds*
Location–shift	100	96.2	3.8	0	0
	200	99.6	0.4	0	0
	400	100	0	0	0
Accelerated failure time and proportional hazard	100	0.8	43.4	35.6	20.2
	200	0.2	39	43.8	17
	400	0	47.8	44.2	8

†The first column lists the true covariate model; the second column lists the sample size; the last four columns contain the proportion of each of the four covariate models selected.

Table 4. Finite sample performance of $\hat{\theta}'$†

Model	*Results for the following models:*			
	Location–shift	*Accelerated failure time*	*Proportional hazard*	*Proportional odds*
Clayton	−0.0024 (0.1113)	0.0029 (0.1736)	−0.0022 (0.1507)	0.0039 (0.2683)
	−0.0015 (0.1105)	0.0058 (0.1662)	−0.0032 (0.1514)	−0.0023 (0.2633)
Frank	−0.0042 (0.1016)	−0.0084 (0.1734)	0.0067 (0.1544)	−0.0028 (0.2573)
	0.0011 (0.0995)	−0.0094 (0.1661)	−0.0096 (0.1680)	0.0013 (0.2602)

†The first number in each column is the average bias of $\hat{\theta}_1$, the second number in parentheses is the standard deviation of $\hat{\theta}_1$ based on 1000 replications, the third number is the average bias of $\hat{\theta}_2$ and the fourth number in parentheses is the standard deviation of $\hat{\theta}_2$ based on 1000 replications.

larger. When the true model changes to the accelerated failure time and proportional hazard models, which are both correct in our setting, these two models together are chosen most of the time. As n increases, the proportion of a correct decision also increases (79% when $n=100$, 82.8% when $n=200$ and 92% when $n=400$).

We also examined the situation of multiple covariates. With $Z'=(Z^{(1)}, Z^{(2)})$ in which $Z^{(j)}$ $(j=1,2)$ are both binary, the sample can be portioned into four groups with $Z_1'=(0,0)$ $(\tau=0.2)$, $Z_2'=(0,1)$ $(\tau=0.3)$, $Z_3'=(1,0)$ $(\tau=0.4)$ and $Z_4'=(1,1)$ $(\tau=0.5)$. The sample size in each of the four groups is 75. We evaluated two dependence structures, namely Clayton and Frank, and four regression models, namely location–shift, accelerated failure time, proportional hazard and proportional odds, for each of which $\theta_0'=(0.3,0.3)$. The average bias and the standard deviation on the basis of 1000 simulation runs are reported in Table 4. The results show that the method proposed still performs well under various regression settings.

4.2. Data analysis

The methodology proposed is applied to the bone marrow transplants data that were given in Klein and Moeschberger (2003), page 484. There were 137 leukaemia patients receiving bone marrow transplants. Let X be the time to relapse of leukaemia, Y be the time to death and C

be the time from transplant to the end of study. Let $\delta_x = I(X \leqslant Y \wedge C)$ be the relapse indicator and let $\delta_y = I(Y \leqslant C)$ be the death indicator. The sample can be divided into three groups with $Z' = (0,0)$ indicating the acute myelogenous leukaemia (AML) low risk group, $Z' = (0,1)$ indicating the acute lymphoblastic leukaemia (ALL) group and $Z' = (1,0)$ indicating the AML high risk group. The regression model of interest is $h(X) = -Z'\theta + \varepsilon$, where $\theta' = (\theta_1, \theta_2)$ which measures whether the disease type affects the relapse time.

For each covariate group, we test the hypothesis $H_0 : \phi_\alpha \sim$ Clayton model *versus* H_a: not H_0. By bootstrapping 1000 times, the p-values of D^C for the AML high risk group, the ALL group and the AML low risk group are 0.752, 0.656 and 0.177 respectively. Hence the Clayton model is adopted for all three groups. Using Day's method (or equivalently Wang's method) to estimate τ_z, we obtain $\hat{\tau}_{(0,0)} = 0.7485$ (0.1176), $\hat{\tau}_{(0,1)} = 0.7894$ (0.0853) and $\hat{\tau}_{(1,0)} = 0.7685$ (0.0872), where the number in parentheses is the estimated standard derivation by using the jackknife method. The above analysis implies that the dependence structures in the three groups are similar and the two events are highly correlated.

Then we choose a model for measuring the group effect on X. Fig. 1 shows the fitted log–log-plot of $\hat{F}_x(x)$ for the three groups. Since the three curves look parallel, we choose the proportional hazard model to measure the group effect. On the basis of the method that was described in Section 3.2, we can formally test the proportional hazard model assumption. By bootstrapping 1000 times, we obtain p-value 0.774 which implies that this model is appropriate. Fig. 2 depicts the three survival curves of $\hat{F}_x(x)$. Under the proportional hazard regression model and the Clayton assumption for each covariate group, we obtain $\hat{\theta}_1 = 1.3624$ (0.3765) and $\hat{\theta}_2 = 0.9503$ (0.3984). The results show that the risk of relapse for the AML high risk group is 3.9 times that of the risk for the AML low risk group, and the risk for the ALL group is 2.59 times that of the AML low risk group. The difference is statistically significant.

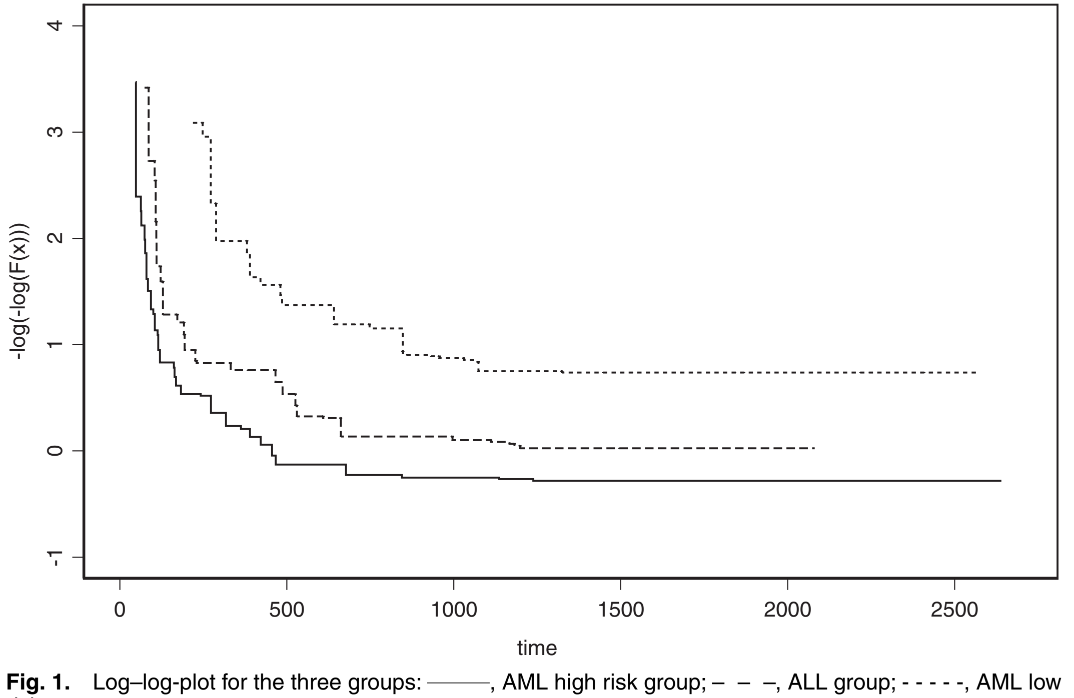

Fig. 1. Log–log-plot for the three groups: ———, AML high risk group; – – –, ALL group; - - - - -, AML low risk group

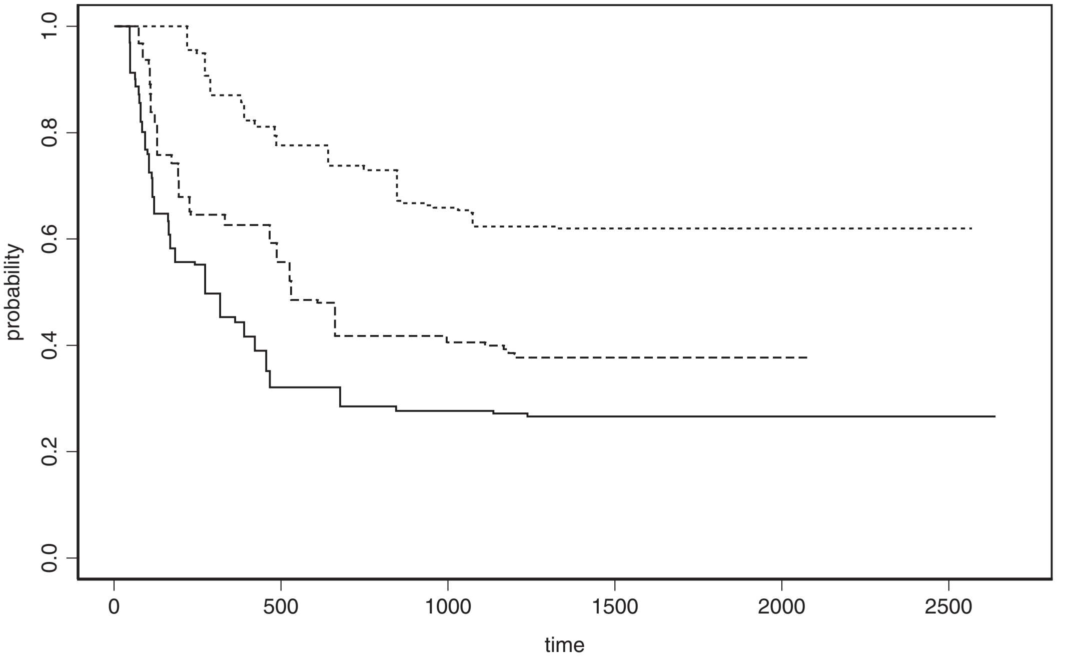

Fig. 2. $\hat{F}_X(t)$ for the three groups: ———, AML high risk group; – – –, ALL group; - - - - -, AML low risk group

5. Concluding remarks

In this paper, we model the failure time to a non-terminal event by a flexible transformation model. To handle the problem of dependent censoring, we make an additional assumption that, for each covariate group, failure times for the two types of event follow a copula model in the identifiable region. Model checking procedures are also proposed to examine the appropriateness of these two model assumptions. Compared with existing methods such as that proposed by Lin *et al.* (1996), our approach allows for different dependence structures in each group, avoids making additional modelling assumptions on the terminal event and utilizes all the data without paying the price for artificial censoring. The simulation analysis confirms our conjecture that the estimator that was proposed by Lin *et al.* (1996) becomes unreliable if the dependence structures in the two groups are different.

The strategy proposed for checking the copula assumption is to compare the non-parametric estimator with its model-based estimator of a chosen reference function, say $F^{11}(t_1, t_2)$. This technique is similar to that used in Wang and Wells (2000). For possible future research, one may examine how to choose such a function or a combination of several functions that contain most of the model information that is characterized by $\phi(\cdot)$ so that the corresponding test procedure would detect the departure from the null hypothesis better and hence give higher power. To select an appropriate regression model for the non-terminal event, a formal model checking procedure is also proposed by using the bootstrap method. The regression method proposed can handle multiple covariates with discrete values. Extension to continuous covariates must face the challenge of imposing additional regression assumptions on model (2) or adopting some non-parametric techniques like smoothing. This goes beyond the scope of the current paper but may deserve further investigation. Note that in model (3) we suggest use of the Kaplan–Meier estimator based on data $\{(\tilde{Y}_i, 1-\delta_{yi})\ (i=1,\ldots,n)\}$ to estimate $G(t)$. Since C is also censored by $X \wedge Y$, another estimator based on data $\{(\tilde{X}_i, 1-\delta_{xi}\delta_{xi})\ (i=1,\ldots,n)\}$ can be constructed. Obviously the latter yields a worse estimator of $G(t)$. However, it has been shown in other contexts, such as in Tsai and Crowley (1998), that plugging in a worse estimator of the nuisance parameter sometimes improves the result.

Acknowledgements

The first two authors were funded by the National Science Council of Taiwan. Part of the work was done while the second author was visiting the Institute for Mathematical Sciences, National University of Singapore, in 2005. The visit was supported by the Institute.

Appendix A: Regularity conditions of theorem 1

Let

$$\bar{U}(\theta) = \sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta) z_{kj} \sqrt{\left(\frac{c_k c_j}{c_k + c_j}\right)} \int_0^{T_{kj}} W_{kj}(t)\, g_{kj}(t,\theta)\, \mathrm{d}t \qquad (12)$$

where $c_k = \lim_{n\to\infty}(n_k/n)$, and $(0, T_{kj})$ is the support of $\tilde{X}$ in the subgroup with $Z = z_k$ or $Z = z_j$. To derive large sample properties of θ, we assume the following regularity conditions.

(a) As $n\to\infty$, $c_k = \lim_{n\to\infty}(n_k/n) > 0$ for all k values.
(b) For each $Z = z_k$, the $H_{z_k}(u, v, \alpha)$ has bounded partial derivatives with respect to u, v and α, where $H_z(u, v, \alpha) = \phi_{z,\alpha}^{-1}\{\phi_{z,\alpha}(u) + \phi_{z,\alpha}(v)\}$ is defined in model (2).
(c) For each $Z = z_k$, the standard regularity conditions hold for estimating $F_{z_k}(x, x)$ and $F_{y,z_k}(x)$ (e.g. conditions for theorem 6.3.2 in Fleming and Harrington (1991)) so that ${n_k}^{1/2}\{\hat{F}_{z_k}(x,x) - F_{z_k}(x,x)\}$ and ${n_k}^{1/2}\{\hat{F}_{y,z_k}(x) - F_{y,z_k}(x)\}$ converge weakly to Gaussian processes.
(d) The weight functions $w_0(x)$ and $W_{kj}(t)$ are positive and bounded and $w_0(x)$ is differentiable with continuous derivatives.
(e) For each of the two classes of model (1), we impose the following assumptions:
   (i) for the first case, $h(t)$ is differentiable, $h'(t) \neq 0$ and is continuous, $\tilde{W}_{kj}(t) = W_{kj}(t)/h'(t)$ is differentiable and $\int |\tilde{W}_{kj}(t)|\, \mathrm{d}t < \infty$;
   (ii) for the second case, the distribution of ε has a density $f_\varepsilon(t)$ which is differentiable with bounded derivatives.
(f) The function $\bar{U}(\theta)$ which is defined in equation (12) is differentiable with respect to θ and the matrix

$$\left(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0), \ldots, \frac{\partial}{\partial\theta_p}\bar{U}(\theta_0)\right)$$

is non-singular. Furthermore $\bar{U}(\theta) \neq 0$ for $\theta \neq \theta_0$ and $\liminf_{\|\theta\|\to\infty}|\bar{U}(\theta)| > 0$.

Appendix B: Sketch of proof for theorem 1

Here we provide a brief sketch of the proof of theorem 1; the details are given in Hsieh *et al.* (2007).
Consider

$$U(\theta) = \sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta) z_{kj} \sqrt{\left\{\frac{(n_k/n) n_j/n}{n_k/n + n_j/n}\right\}} \int_0^{t_{kj}} W_{kj}(t)\, \hat{g}_{kj}(t,\theta)\, \mathrm{d}t,$$

where $W_{kj}(t)$ is a deterministic function. Equation (A.5) in Hsieh *et al.* (2007) states that $U(\theta)/n^{1/2}$ converges (in probability) to

$$\bar{U}(\theta) = \sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta) z_{kj} \sqrt{\left(\frac{c_k c_j}{c_k + c_j}\right)} \int_0^{T_{kj}} W_{kj}(t)\, g_{kj}(t,\theta)\, \mathrm{d}t,$$

in which the convergence is uniform in θ. Consider a compact set $D_r=\{\|\theta-\theta_0\|\leqslant r\}$ where r is a positive constant. By assumption (f) $\bar{U}(\theta)\neq 0$ for $\theta\neq\theta_0$ and $\liminf_{\|\theta\|\to\infty}|\bar{U}(\theta)|>0$; then the continuity of $\bar{U}(\theta)$ implies that $\inf_{\|\theta-\theta_0\|>r}|\bar{U}(\theta)|>0$. The (uniform) convergence of $U(\theta)/n^{1/2}$ to $\bar{U}(\theta)$ implies that there will be no solution for $U(\theta)=0$ outside the compact set D_r when n is large. Since this is true for every $r>0$, $\hat{\theta}$ is consistent.

By Taylor series expansion we obtain

$$U(\hat{\theta})=0=U(\theta_0)+\sum_{l=1}^{p}\frac{\partial}{\partial\theta_l}U(\check{\theta})(\hat{\theta}_l-\theta_{l,0}), \tag{13}$$

where $\check{\theta}$ is an intermediate value between $\theta_0=(\theta_{1,0},\ldots,\theta_{p,0})^{\mathrm{T}}$ and $\hat{\theta}=(\hat{\theta}_1,\ldots,\hat{\theta}_p)^{\mathrm{T}}$. Hence we have the expression

$$\frac{1}{n^{1/2}}\left(\frac{\partial}{\partial\theta_1}U(\check{\theta}),\ldots,\frac{\partial}{\partial\theta_p}U(\check{\theta})\right)n^{1/2}(\hat{\theta}-\theta_0)=-U(\theta_0). \tag{14}$$

The statement of expression (A.6) in Hsieh *et al.* (2007) is about the convergence of

$$\frac{1}{n^{1/2}}\frac{\partial}{\partial\theta_l}U(\theta)$$

to

$$\frac{\partial}{\partial\theta_l}\bar{U}(\theta)$$

locally uniformly at $\theta=\theta_0$. Using this condition along with the consistency of $\hat{\theta}$, we can show that

$$\frac{1}{n^{1/2}}\left(\frac{\partial}{\partial\theta_1}U(\check{\theta}),\ldots,\frac{\partial}{\partial\theta_p}U(\check{\theta})\right)\xrightarrow{\mathrm{P}}\left(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0),\ldots,\frac{\partial}{\partial\theta_p}\bar{U}(\theta_0)\right)$$

which, by assumption (f), is a non-singular constant matrix. By expression (A.7) in Hsieh *et al.* (2007), $U(\theta_0)$ is asymptotic normal with mean 0. Therefore $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotic normal with mean 0 because it has the same asymptotic distribution as

$$-\left(\frac{\partial}{\partial\theta_1}\bar{U}(\theta_0),\ldots,\frac{\partial}{\partial\theta_p}\bar{U}(\theta_0)\right)^{-1}U(\theta_0).$$

This completes the proof.

Appendix C: Sketch of proof for theorem 2

Compared with the previous proof, we need to show only that

(a) $\{\hat{U}(\theta)-U(\theta)\}/n^{1/2}$ uniformly strongly converges to 0,
(b) $\partial[\{\hat{U}(\theta)-U(\theta)\}/n^{1/2}]\partial\theta_l$ strongly converges to 0, which takes place locally uniformly at $\theta=\theta_0$, and
(c) $\hat{U}(\theta_0)-U(\theta_0)$ strongly converges to zero.

Firstly,

$$\frac{\hat{U}(\theta)-U(\theta)}{n^{1/2}}=\sum_{k<j}w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t. \tag{15}$$

Under the related regularity conditions and the uniform and strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$, we can establish the uniform and strong convergence of expression (15) to 0. So condition (a) holds.

Secondly, we can write

$$\hat{U}(\theta_0)-U(\theta_0)=\sum_{k<j}w_0(z_{kj}^{\mathrm{T}}\theta_0)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,n^{1/2}\,\hat{g}_{kj}(t,\theta_0)\,\mathrm{d}t. \tag{16}$$

Under the related regularity conditions as well as $n^{1/2}\,\hat{g}_{kj}(t,\theta_0)=O_p(1)$ for all t and $\hat{W}_{kj}(t)-W_{kj}(t)=o_p(1)$, we can show strong convergence of expression (16) to 0. So condition (c) holds.

Finally

$$\frac{\partial}{\partial\theta_l}\left\{\frac{\hat{U}(\theta)-U(\theta)}{n^{1/2}}\right\}=\sum_{k<j} w_0'(z_{kj}^{\mathrm{T}}\theta)z_{kj}z_{kj,l}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t$$
$$+\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\frac{\partial}{\partial\theta_l}\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t. \qquad (17)$$

The required regularity conditions plus the uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$ imply that the first term in equation (17) converges uniformly (in θ) and strongly to 0. To prove the second term, we need to consider the two regression classes separately.

For the first case that $\xi_\theta(F)(t)=F[h^{-1}\{h(t)+\theta\}]$,

$$\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\frac{\partial}{\partial\theta_l}\left[\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t\right]$$
$$=\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}$$
$$\times\int_{h^{-1}\{h(0)+z_{kj}^{\mathrm{T}}\theta\}}^{h^{-1}\{h(t_{kj})+z_{kj}^{\mathrm{T}}\theta\}}\frac{(\hat{W}_{kj}[h^{-1}\{h(t*)-z_{kj}^{\mathrm{T}}\theta\}]-W_{kj}[h^{-1}\{h(t*)-z_{kj}^{\mathrm{T}}\theta\}])z_{kj,l}}{h'[h^{-1}\{h(t*)-z_{kj}^{\mathrm{T}}\theta\}]}\,\mathrm{d}\hat{F}_{x,z_k}(t*).$$

(See the proof of expression (A.4) in Hsieh *et al.* (2007).) Note that local boundedness of $1/h'[h^{-1}\{h(t*)-z_{kj}^{\mathrm{T}}\theta\}]$ can be established owing to the continuity of $h'(t)$. The related regularity conditions and the result of expression (A.4) in Hsieh *et al.* (2007) imply that the second term in equation (17) locally uniformly strongly converges to 0.

For the second case that $\xi_\theta(F)(t)=F_\varepsilon[F_\varepsilon^{-1}\{F(t)\}+\theta]$, then $\xi_\theta'(F)(t)=f_\varepsilon[F_\varepsilon^{-1}\{F(t)\}+\theta]$. Hence

$$\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\frac{\partial}{\partial\theta_l}\left[\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,\hat{g}_{kj}(t,\theta)\,\mathrm{d}t\right]$$
$$=\sum_{k<j} w_0(z_{kj}^{\mathrm{T}}\theta)z_{kj}\sqrt{\left\{\frac{(n_k/n)n_j/n}{n_k/n+n_j/n}\right\}}\int_0^{t_{kj}}\{\hat{W}_{kj}(t)-W_{kj}(t)\}\,z_{kj,l}\,f_\varepsilon[F_\varepsilon^{-1}\{\hat{F}(t)\}+z_{kj}^{\mathrm{T}}\theta]\,\mathrm{d}t,$$

which converges uniformly to 0, owing to the related regularity conditions, uniform strong convergence of $\hat{W}_{kj}(t)$ to $W_{kj}(t)$ and the boundedness of $f_\varepsilon[F_\varepsilon^{-1}\{\hat{F}(t)\}+z_{kj}^{\mathrm{T}}\theta]$, i.e. the second term in equation (17) locally uniformly strongly converges to 0.

In summary equation (17) locally uniformly strongly converges to 0, i.e. condition (b) holds. This completes the proof.

Appendix D: Sketch of proof for theorem 3

First, the empirical distribution function is uniformly consistent, i.e. $\sup_{t_1\leqslant t_2}|\hat{F}^{11}(t_1,t_2)-F^{11}(t_1,t_2)|\to^{\mathrm{P}}0$ as $n\to\infty$. Then it can be shown that, for $t_1\leqslant t_2$, $\tilde{F}_k^{11}(t_1,t_2)$ uniformly converges to

$$\bar{F}_k^{11}(t_1,t_2,\hat{\alpha})=\frac{\int_{y=t_2}^{\infty}\int_{x=t_1}^{y}\bar{F}_k(\mathrm{d}x,\mathrm{d}y,\hat{\alpha})\,G(y)}{\int_{y=0}^{\infty}\int_{x=0}^{y}\bar{F}_k(\mathrm{d}x,\mathrm{d}y,\hat{\alpha})\,G(y)},$$

where $\bar{F}_k(\mathrm{d}x,\mathrm{d}y,\hat{\alpha})=\bar{F}_k(x,y,\hat{\alpha})-\bar{F}_k(x+\mathrm{d}x,y,\hat{\alpha})-\bar{F}_k(x,y+\mathrm{d}y,\hat{\alpha})+\bar{F}_k(x+\mathrm{d}x,y+\mathrm{d}y,\hat{\alpha})$ and $\bar{F}_k(x,y,\hat{\alpha})=\phi_{\hat{\alpha}}^{(k)^{-1}}[\phi_{\hat{\alpha}}^{(k)}\{F_x(x)\}+\phi_{\hat{\alpha}}^{(k)}\{F_y(y)\}]$. If $\phi_\alpha^{(k)}$ is the true copula model, then $\hat{\alpha}\to^{\mathrm{P}}\alpha$ and $\bar{F}_k^{11}(t_1,t_2,\hat{\alpha})$ uniformly converges to $F^{11}(t_1,t_2)$. Therefore, $D^k\to^{\mathrm{P}}0$.

If $\phi_\alpha^{(k)}$ is not the true model, let $d_k(\alpha^*) = \sup_{t_1 \leqslant t_2} |\bar{F}_k^{11}(t_1, t_2, \alpha^*) - F^{11}(t_1, t_2)|$. We can show that $d_k(\alpha^*) > 0$ for all $\alpha^* \in A_k$. To see this, let us consider when $d_k(\alpha^*) = 0$, i.e.

$$\frac{\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} \bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*)\, G(y)}{\int_{y=0}^{\infty} \int_{x=0}^{y} \bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*)\, G(y)} = \frac{\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} F(\mathrm{d}x, \mathrm{d}y)\, G(y)}{\int_{y=0}^{\infty} \int_{x=0}^{y} F(\mathrm{d}x, \mathrm{d}y)\, G(y)}$$

for all $t_1 \leqslant t_2$, where $F(\mathrm{d}x, \mathrm{d}y) = F(x, y) - F(x + \mathrm{d}x, y) - F(x, y + \mathrm{d}y) + F(x + \mathrm{d}x, y + \mathrm{d}y)$ and $F(x, y) = {\phi_\alpha}^{-1}[\phi_\alpha\{F_x(x)\} + \phi_\alpha\{F_y(y)\}]$. Let

$$p^* = \int_{y=0}^{\infty} \int_{x=0}^{y} \bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*)\, G(y)$$

and

$$p = \int_{y=0}^{\infty} \int_{x=0}^{y} F(\mathrm{d}x, \mathrm{d}y)\, G(y).$$

Note that p and p^* are constants independent of t_1 and t_2. Hence the above equation becomes

$$\int_{y=t_2}^{\infty} \int_{x=t_1}^{y} \left\{ \frac{\bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*)}{p^*} - \frac{F(\mathrm{d}x, \mathrm{d}y)}{p} \right\} G(y) = 0$$

for all $t_1 \leqslant t_2$. Therefore,

$$\bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*)/p^* - F(\mathrm{d}x, \mathrm{d}y)/p = 0$$

on the region $\{(x, y) : x \leqslant y \text{ and } G(y) > 0\}$. Consider the variables $u = F_x(x)$ and $v = F_y(y)$; it is easy to see that $F(\mathrm{d}x, \mathrm{d}y) = C_\alpha(\mathrm{d}u, \mathrm{d}v)$ and $\bar{F}_k(\mathrm{d}x, \mathrm{d}y, \alpha^*) = C_{\alpha^*}^{(k)}(\mathrm{d}u, \mathrm{d}v)$. By the assumption that C has larger support than the supports of X and Y, both $F_x(x)$ and $F_y(y)$ change from 1 to 0 on the region $\{(x, y) : x \leqslant y \text{ and } G(y) > 0\}$. Therefore we have

$$C_{\alpha^*}^{(k)}(\mathrm{d}u, \mathrm{d}v)/p^* - C_\alpha(\mathrm{d}u, \mathrm{d}v)/p = 0$$

on the region $\{(u, v) : 0 \leqslant F_x\{F_y^{-1}(v)\} \leqslant u \leqslant 1\}$. Therefore, the analytical properties of $\phi_{\alpha^*}^{(k)}$ and ϕ_α imply that the above equality holds over the region $\{(u, v) : 0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1\}$. This together with the fact that $C_{\alpha^*}^{(k)}(0, 0) = C_\alpha(0, 0) = 0$ imply that

$$p\, C_{\alpha^*}^{(k)}(u, v) = p^*\, C_\alpha(u, v)$$

on the region $\{(u, v) : 0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1\}$. Since $C_{\alpha^*}^{(k)}(1, 1) = C_\alpha(1, 1) = 1$, $p = p^*$. Now, $C_{\alpha^*}^{(k)}(u, v) = C_\alpha(u, v)$ on the region $\{(u, v) : 0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1\}$. This contradicts the fact that $\phi_\alpha^{(k)}$ is not the true copula model. Hence $d_k(\alpha^*) > 0$ for all $\alpha^* \in A_k$. This fact together with the closedness of A_k and the continuity in α^* imply that $d_k = \inf_{\alpha^* \in A_k} \{d_k(\alpha^*)\} > 0$. Therefore, if model k is wrong, $D^k \rightarrow^{\mathrm{P}} d_k > 0$.

Suppose that there are K candidate models under consideration. Let

$$d = \inf_{\{k : 1 \leqslant k \leqslant K,\, \phi_\alpha^{(k)} \text{ is not true copula model}\}} (d_k).$$

Then $d > 0$. And, as $n \rightarrow \infty$, $\Pr(D^k > d/2) \rightarrow 1$ if model k is wrong whereas $\Pr(D^k < d/2) \rightarrow 1$ if model k is correct. Therefore $\hat{k}$ is consistent.

References

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.

Day, R., Bryant, J. and Lefkopoulou, 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.

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

Genest, C. (1987) Frank's family of bivariate distributions. *Biometrika*, **74**, 549–555.
Ghosh, D. and Lin, D. Y. (2003) Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, **59**, 877–885.
Gumbel, E. J. (1960) Bivariate exponential distributions. *J. Am. Statist. Ass.*, **55**, 698–707.
Hsieh, J. J., Wang, W. and Ding, A. (2007) Regression analysis based on semi-competing risks data. *Technical Report*. Institute of Statistics, National Chiao-Tung University, Hsin-Chu. (Available from `http://www.math.neu.edu/ding/paper/JRSSB07/techreport.pdf`.)
Klein, J. P. and Moeschberger, M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edn. New York: Springer.
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.
Lin, D. Y. and Ying, Z. (2003) Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics*, **4**, 385–398.
Prentice, R. L. and Cai, J. (1992) Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495–512.
Tsai, W.-Y. and Crowley, J. (1998) A note on nonparametric estimators of the bivariate survival function under univariate censoring. *Biometrika*, **85**, 573–580.
Wang, W. (2003) Estimating the association parameter for copula models under dependent censoring. *J. R. Statist. Soc.* B, **65**, 257–273.
Wang, W. and Wells, M. T. (2000) Model selection and semiparametric inference for bivariate failure-time data. *J. Am. Statist. Ass.*, **95**, 62–72.