Biometrika (1996), **83**, 1, *pp.* 29–39
Printed in Great Britain

Modelling failure-time associations in data with multiple levels of clustering

By KAREN J. BANDEEN-ROCHE AND KUNG-YEE LIANG
Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205, U.S.A.

Summary

In recent years substantial research has been devoted to developing failure-time methodology which accounts for possible dependency between observations. An example is the univariate frailty model (Vaupel, Manton & Stallard, 1979), which incorporates an exchangeable dependence structure by the inclusion of cluster-specific random effects. In some studies it may be reasonable to expect more than one level of within-cluster association: for instance, association between a parent and child versus that between two siblings in studies of familial disease aggregation, or association between two village residents who live in different households versus that between residents of the same household in intervention studies. We propose a family of distributional models for failure-time data that accounts for multiple levels of clustering and reduces in the case of a single clustering level to a univariate frailty model. The resulting survival functions are constructed by a recursive nesting of univariate frailty-type distributions through which archimedean copula forms are determined for all bivariate margins. Properties of the proposed model are developed, illustrated and briefly contrasted with multivariate frailty model properties. In conclusion, we outline the application of our model to marginal risk regression problems.

Some key words: Archimedean copula; Conditional hazard ratio; Frailty distribution; Hierarchical; Laplace transform; Multivariate survival function.

1. Introduction

In recent years a growing body of work has appeared on the analysis of failure-time data when outcomes are not all independent. Lack of independence may occur when there are multiple potential failure events to be studied per unit or when units cluster. When $m > 2$ outcomes are to be observed per cluster, some of the pairwise associations among cluster members may be stronger than others. In this case, we say that these are multiple levels of clustering. For motivation, consider two such cases. First, understanding the strength of intra-familial association provides insight into the role of heredity in chronic disease onset; multiple-level modelling is indicated because, for instance, parent–parent association may differ in strength from sibling–sibling association. Secondly, understanding strength of association within geopolitical units assists public health planners in developing intervention strategies by indicating appropriate units for health care delivery; multiple-level modelling is indicated because within-household association might well exceed within-village, between-household association.

In this paper we propose a family of distributional models for multivariate survival data which allows for multiple levels of association. Generated by building specifiable distributional forms within each other recursively, it generalises ideas previously discussed by Joe (1993). Because its construction induces an ordering on pairwise associations, this latter family is particularly appropriate for geopolitical-type data with a hierarchical structure. Two particularly appealing features of the construction are that it permits specification of simple distributional forms for all bivariate margins and that the resulting distribution reduces in the case of a single clustering level to the frailty formulation introduced by Vaupel et al. (1979).

In the next section frailty models are reviewed. Section 3 is devoted to developing our proposed model; conditions under which that model represents a legitimate survival function comprise the chief result of the paper. An example which illustrates strengths and limitations of the approach is presented in § 4. We conclude with a discussion of regression applications and possibilities for future research.

2. Frailty models

For simplicity, consider a single cluster of size m. Let T_j denote the survival time for individual j in the cluster, $j=1,\ldots,m$; $S_j(t):=\mathrm{pr}(T_j>t)$, the marginal survival function for the jth individual; and $S(t):=\mathrm{pr}(T_1>t_1,\ldots,T_m>t_m)$, the joint survival function for the cluster, with t denoting the vector $(t_1,\ldots,t_m)'$. Further, let $\lambda_j(t)$ denote the hazard function corresponding to S_j.

The univariate frailty model is a random-effects formulation for within-cluster association; here we recapitulate an excellent summary provided by Oakes (1989). According to this model, association is generated because each cluster carries a corresponding random effect, denoted by α, with distribution G and Laplace transform $p(x):=E(e^{-x\alpha})$. Conditional upon α cluster survival times are assumed to be independent with survival functions $\mathrm{pr}(T_j>t\mid\alpha=a)=\{S_j^*(t)\}^a$, for some continuous survival functions S_j^* $(j=1,\ldots,m)$. Because $S_j(t)=\int\{S_j^*(t)\}^a\,dG(a)$, the multivariate survival function for the cluster may be specified as a function of its marginals as follows:

$$S(t)=\int\exp\left\{a\sum_{j=1}^{m}\log S_j^*(t_j)\right\}dG(a)=p\left\{-\sum_{j=1}^{m}\log S_j^*(t_j)\right\}=p\left[\sum_{j=1}^{m}q\{S_j(t_j)\}\right],\quad(1)$$

where q is the inverse function of p. This last functional relationship is special because it defines a proper subclass of the symmetric multivariate distributions with uniform marginals which are archimedean copulas, described by Genest & MacKay (1986a, b). For such archimedean copula distributions, the association between T_j and T_k $(j<k=2,\ldots,m)$ may be described by the commonly-used conditional hazard ratio

$$\theta(t)=\theta(t_j,t_k):=\frac{\lambda_{T_j\mid T_k}(t_j\mid T_k=t_k)}{\lambda_{T_j\mid T_k}(t_j\mid T_k>t_k)}.$$

This measure is easily interpretable as the factor by which an individual's hazard at time t_j is increased if his cluster partner is known to have failed at time t_k, rather than to have survived past t_k. One famous example of (1) is the Clayton (1978) model, which is characterised by constant hazard ratio $\theta(t)\equiv\theta$ corresponding to gamma frailty with $p(u)=(1+u)^{1/(1-\theta)}$ $(1\leqslant\theta<\infty)$. Substituting into (1), this leads to

$$S(t)=\left[\sum_{j=1}^{m}\{S_j(t_j)\}^{1-\theta}-m+1\right]^{1/(1-\theta)}.\quad(2)$$

Marshall & Olkin (1988), Hutchinson (1981), A. Yashin, J. Vaupel and I. Iachine, in the unpublished report 'Correlated individual frailty: an advantageous approach to survival analysis of bivariate data', and others have described generalisations of (1) which permit individual cluster members to carry different, possibly dependent random effects. A special case of such multivariate frailty models is particularly natural for applications of present interest, as follows. Consider a cluster which may be partitioned into N subclusters $I_1, \ldots, I_N$ such as siblings versus each parent or households within a village. Then, a reasonable multivariate survival function for the cluster is given by

$$S(t) = \int \exp\left[-\sum_{k=1}^{N} a_k \sum_{j \in I_k} q_k\{S_j(t_j)\}\right] dG(a_1, \ldots, a_N). \tag{3}$$

Two properties of this formulation have motivated our development of the alternative multivariate survival model to be described in the next section. First, whereas the multivariate survival function for members of any single subcluster has the archimedean copula form of equation (1), the bivariate survival function for members of distinct subclusters takes a more general form which does not generally define an archimedean copula. Therefore, for convenient multivariate frailty distribution choices, between- and within-subcluster associations θ may differ substantially in complexity of time dependence. Secondly, the continuing development of multivariate frailty distributions notwithstanding, the diversity and computational tractability of existing univariate frailty distributions remains appealing. Thus, we proceed to develop an alternative family of multilevel survival functions whose construction is based exclusively on univariate frailty distributions.

3. A fully hierarchical model proposal

3·1. *Notational framework*

The basic idea of our proposal is to build the cluster survival function $S(t)$ recursively from archimedean copula distributions such as (1). At all but the first stage of the recursion, however, the survival function arguments of the copula may themselves be multivariate.

For the sake of full generality, the proposed model relies on a somewhat cumbersome notational framework based on nested partitioning of cluster members. This partitioning defines a series of subclusters within the overall cluster. Heuristically, at each 'level' in the series we aggregate subclusters from the previous level; increasing levels then correspond to increasingly dissociated subclusters. Let l be an index which tracks levels, $l = 0, \ldots, L$. At $l = 0$ cluster members are considered as individuals; at $l = 1$ individuals are grouped into initial subclusters, such as households; increasing values of l denote increasingly dissociated units, such as villages at $l = 2$, regions at $l = 3$, etc. The index k will track the number of cells in the partition per level, $k = 1, \ldots, N_l$. Then cluster partitions are defined by the index sets

$$I_{lk} = \{j : \text{cluster member } j \in k\text{th cell partition at level } l\};$$

these sets are constructed to satisfy

$$I_{lk} \cap I_{lr} = \varnothing \quad (k \neq r), \quad \bigcup_{k=1}^{N_l} I_{lk} = \{1, \ldots, m\}$$

for all l. Finally, let M_{lk} identify the cells of the $(l-1)$th partition which are collapsed to create the kth cell of the lth partition; $M_{lk} = \{r : I_{lk} = \bigcup_r I_{l-1,r}\}$;

$$M_{lk} \cap M_{lr} = \varnothing \quad (k \neq r), \quad \bigcup_{k=1}^{N_l} M_{lk} = \{1, \ldots, N_{l-1}\}.$$

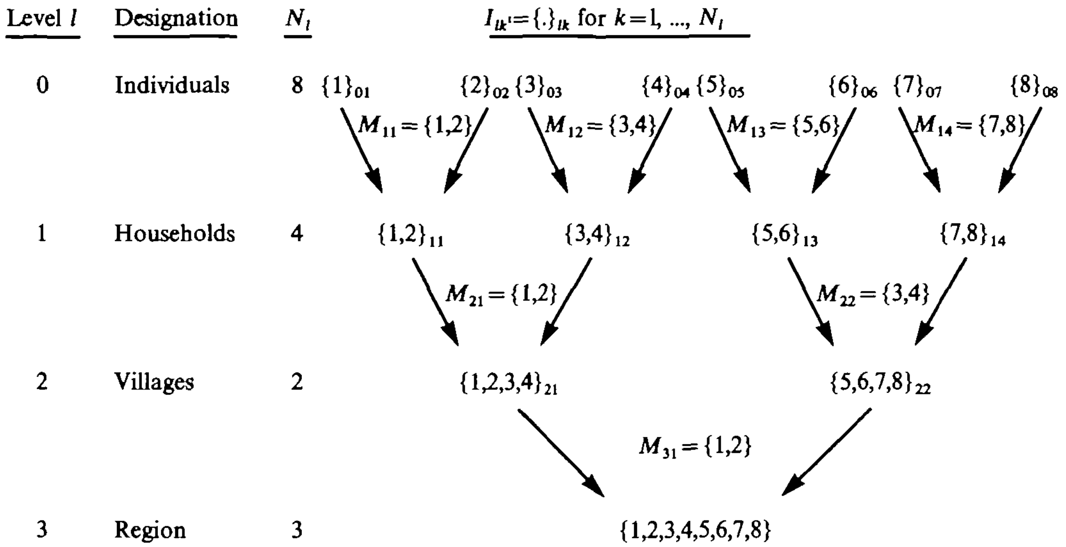

Fig. 1. Hierarchical notation for an eight-member geographical cluster with $L=3$.

Ultimately a single partition cell will engulf all indices, so that $N_L = 1$. This framework is illustrated in Fig. 1. In terms of the conditional hazard ratio, the association between persons j and j' is

$$\lambda_{T_j|T_{j'}}(t_j|T_{j'}=t_{j'})/\lambda_{T_j|T_{j'}}(t_j|T_{j'}>t_{j'})=:\theta_{rs}(t),$$

where

$$r:=\min\{l: j\in I_{lu} \text{ and } j'\in I_{lu} \text{ for some } u\},\quad s:=\{k: j\in I_{rk}, j'\in I_{rk}\},\quad t:=(t_j, t_{j'})'.$$

When modelling multiple clusters jointly, we permit clusters of different sizes. Thus, distinct clusters $T_a=(T_{a1}, \ldots, T_{am_a})$ and $T_b=(T_{b1}, \ldots, T_{bm_b})$ may differ in their number of levels L_a and L_b and in composition of sublevels $\{I_{lk}\}$ within a given level l. Importantly, though, we impose the condition that the cluster-specific survival functions (5) to follow all derive from a finite set $\{p_{lk};\ k=1, \ldots, N_l, l=1, \ldots, L\}$. Thus, $L_a \leqslant L$ and

$$\{p_{l_ak};\ k=1, \ldots, N_{l_a}, l_a=1, \ldots, L_a\} \subset \{p_{lk};\ k=1, \ldots, N_l, l=1, \ldots, L\}$$

for each cluster a. To complete the joint model, we impose mutual independence between clusters. Summarising, then, our association model has finite dimension independent of the number of clusters.

3·2. *Model specification*

We begin by considering the case of two clustering levels. For this motivational case, assume five cluster members with $I_{11}=\{1, 2\}$, $I_{12}=\{3, 4, 5\}$ and $I_{21}=\{1, 2, 3, 4, 5\}$. Then, the proposed model is

$$S(t)=p_{21}[q_{21}\{S_{11}(t_1, t_2)\}+q_{21}\{S_{12}(t_3, t_4, t_5)\}], \tag{4}$$

with

$$S_{11}(t_1, t_2)=p_{11}\left[\sum_{j=1}^{2} q_{11}\{S_j(t_j)\}\right],\quad S_{12}(t_3, t_4, t_5)=p_{12}\left[\sum_{j=3}^{5} q_{12}\{S_j(t_j)\}\right].$$

Here, p_{1k} are Laplace transforms for distributions which determine the bivariate association structure within I_{1k} and p_{21} is a Laplace transform which determines the form of association between I_{11} and I_{12}; q_{1k} and q_{21} are the respective inverse functions. It is straightforward to verify that each bivariate marginal follows the archimedean copula structure whereby conditional hazard ratio forms $\theta_{rs}(t)$ are determined. In the two clustering level case this model is equivalent to the multivariate mixture family proposed by Joe (1993), and (4) is essentially equivalent to his equation (4·4).

For an arbitrary number of levels, our model is constructed recursively. For notational convenience, define $S_{lk}(t_{lk}) := \mathrm{pr}(T_j > t_j, j \in I_{lk})$ to be the partially marginalised survival function for the individuals whose indices belong to I_{lk}; thus, t_{lk} is a vector of size $|I_{lk}|$, the number of individuals belonging to I_{lk}. Then,

$$S_{0j}(t_{0j}) = S_j(t_j) \quad (j = 1, \ldots, m), \quad S_{1k}(t_{1k}) = p_{1k}\left[\sum_{j \in I_{1k}} q_{1k}\{S_j(t_j)\}\right] \quad (k = 1, \ldots, N_1), \ldots .$$

In general,

$$S_{lk}(t_{lk}) = p_{lk}\left[\sum_{r \in M_{lk}} q_{lk}\{S_{l-1,r}(t_{l-1,r})\}\right] \quad (k = 1, \ldots, N_l;\ l = 2, \ldots, L). \tag{5}$$

As in the two subcluster case, it is straightforward to verify that each bivariate marginal follows the archimedean copula structure whereby conditional hazard ratio forms $\theta_{rs}(t)$ are determined.

3·3. *Survival function legitimacy*

In this section, we ensure that our proposed model produces legitimate multivariate survival functions. Toward this end we assume the following.

Assumption 1. For each $j = 1, \ldots, m$, $S_j(t)$ is a right-continuous survival function.

Assumption 2. Assume p_{lk} $(l = 1, \ldots, L, k = 1, \ldots, N_l)$ are Laplace transforms of distribution functions whose supports exclude $[-\infty, 0]$ and which have piecewise continuous first derivatives on $(0, \infty)$.

Assumption 3. Assume $S_j(t)$ is differentiable with derivative $-f_j(t)$, for each j.

It follows from standard theory (Churchill, 1972, pp. 46–8) that the p_{lk} and q_{lk} are strictly decreasing elements of $C(\infty)$. Moreover, the ith order derivative of p_{lk} has sign $(-1)^i$ everywhere excluding ∞ for all positive integers i, and the first two derivatives of q_{lk} are respectively negative and positive everywhere excluding 0.

Given Assumptions 1 and 2, it is virtually immediate to demonstrate that the proposed model satisfies four properties of multivariate survival functions. That consequence is embodied in the following statement whose proof is deferred to the Appendix.

PROPOSITION 1. *The multivariate function* $S(t)$ *defined by* (5) *satisfies*

(*i*) $S(u) = 1$ *for all* $u \leqslant 0$,
(*ii*) $\lim_{K \to \infty} \inf_{t > K} S(t) = 0$,
(*iii*) $t_1 < t_2 \Rightarrow S(t_1) \geqslant S(t_2)$,
(*iv*) S *is right continuous.*

Here, all inequalities define componentwise relationships.

To ensure that (5) produces a legitimate survival function, we must still demonstrate that the distribution F corresponding to S assigns nonnegative mass to every region of space. This, in turn, will be true if odd-ordered mixed partial derivatives of S are nonpositive and even-ordered mixed partials are nonnegative almost everywhere. Since this property induces restrictions upon our model, we consider it in detail. For a generic function $h(x)$, let $h^{(i)}$ denote the ith derivative of h with respect to x, with $h' := h^{(1)}$ and $h'' := h^{(2)}$. As we show in the Appendix, a sufficient condition to ensure nonnegative mass assignment almost everywhere is as follows.

THEOREM 1. *If for each level $l < L$ of the recursion in* (5) *and each cell $I_{l-1,r}$ containing at least $|I_{l-1,r}| = 2$ members*

$$(-1)^{(i+1)}(q_{lk} \circ p_{l-1,r})^{(i)}(u) > 0 \tag{6}$$

on $u < \infty$, $i = 1, \ldots, |I_{l-1,r}|$, $r = 1, \ldots, N_{l-1}$, then odd- (even-) ordered mixed partial derivatives of $S(t)$ are nonpositive (nonnegative). In (6), *k is such that $r \in M_{lk}$.*

We organise the preceding results into a formal statement which follows immediately from Proposition 1 and Theorem 1.

COROLLARY. *We have that $S(t) = S_{L1}(t)$ as defined by the recursion in* (5) *is a legitimate survival function provided that Assumptions* 1–3 *and the conditions of Theorem* 1 *hold.*

Briefly, the proof of Theorem 1 relies on a chain rule for higher-order differentials and induction across levels to demonstrate that odd (even)-ordered mixed partials of $q_{lk}\{S_{l-1,r}(t_{l-1,r})\}$ are nonnegative (nonpositive), $l = 2, \ldots, L$, $k = 1, \ldots, N_l$, $r \in M_{lk}$. Two remarks should be made. First, Assumption 3 is not a necessary condition of Theorem 2; it is included here for economy of proof. Secondly, Joe (1993) provides an elegant alternative proof of survival function legitimacy in the two-level case, using a condition which is essentially equivalent to (6) as described in his Condition A, Theorem B, and first paragraph of the derivative for his equations (4·1) and (4·2). Generalisation to arbitrarily many levels, however, is complicated by the element of recursion and seems to require the more direct strategy we have employed.

Hierarchical constraints are induced upon our proposed model by the requirement that no region of space be assigned negative mass embodied in (6). Providing a general interpretation of (6) is difficult; rather, the condition must be verified on a case-by-case basis. However, verification is not difficult in some important cases. An example is provided in the next section.

4. AN EXAMPLE: A GAMMA FRAILTY GENERALISATION

In this section we detail a hypothetical example to illustrate the proposed model. Consider a five-individual village within which members $\{1, 2\}$ and $\{3, 4, 5\}$ comprise separate households; hence, $N_1 = 1$, $I_{11} = \{1, 2\}$ and $I_{12} = \{3, 4, 5\}$. To define the cluster survival function, we set $p_{1k}(u) = (1+u)^{1/(1-\theta_k)}$ $(k = 1, 2)$ and $p_{21}(u) = (1+u)^{1/(1-\theta_3)}$ in (5); note that in this case the constraint $\theta_1 = \theta_2$ might well be reasonable. With these Laplace transform choices, the multivariate survival function as calculated from (4) is

$$S(t) = \left[\left\{\sum_{j=1}^{2} S_j(t_j)^{1-\theta_1} - 1\right\}^{(\theta_3-1)/(\theta_1-1)} + \left\{\sum_{j=3}^{5} S_j(t_j)^{1-\theta_2} - 2\right\}^{(\theta_3-1)/(\theta_2-1)} - 1\right]^{1/(1-\theta_3)}.$$

Clearly, then, all bivariate margins of $S(t)$ have the archimedean copula form (2) with $\theta = \theta_1$ for the members of the first household, $\theta = \theta_2$ for any two members of the second household, and $\theta = \theta_3$ for two members of different households. It follows that all pairwise conditional hazard ratios are constant over time and equal $\theta_k \geqslant 1$ as just described, $k = 1, 2, 3$, with $\theta_k = 1$ corresponding to pairwise independence. By way of contrast, consider a multivariate frailty formulation of the type (3) with $G_{\alpha_1,\alpha_2}(a_1, a_2)$ bivariate gamma, having marginal Laplace transforms $p_k(u) = (1+u)^{1/(1-\theta_k)}$ $(k = 1, 2)$ and $\operatorname{cov}(\alpha_1, \alpha_2) = 1/(\theta_3^* - 1)$. While this model also induces constant within-household hazard ratios, it yields the complicated between-household association of

$$\theta_{12}^*(t_1, t_3) = 1 + \frac{(\theta_3^* - 1)X_1(t_1, \theta_1)X_3(t_3, \theta_2)}{\{X_1(t_1, \theta_1) + D_1\}\{X_3(t_3, \theta_2) + D_2\}},$$

where

$$D_k = (\theta_3^* - \theta_k)\{S_1(t_1)^{1-\theta_1} + S_3(t_3)^{1-\theta_2} - 1\}, \quad X_j(t, \theta) = (\theta - 1)S_j(t)^{1-\theta} \quad (k = 1, 2;\ j = 1, 3).$$

It remains to illustrate the nature of the hierarchical constraint induced by (6), which in the present example involves the functions

$$g_k(u) := q_{21}\{p_{1k}(u)\} = (1+u)^{(1-\theta_3)/(1-\theta_k)} - 1 \quad (k = 1, 2).$$

Now,

$$g_1'(u) = \left(\frac{1-\theta_3}{1-\theta_1}\right)(1+u)^{(\theta_1-\theta_3)/(1-\theta_1)},$$

which is positive assuming $\theta_3 > 1$ and $\theta_1 > 1$. Further,

$$g_1''(u) = \left(\frac{\theta_1-\theta_3}{1-\theta_1}\right)\left(\frac{1-\theta_3}{1-\theta_1}\right)(1+u)^{(2\theta_1-\theta_3-1)/(1-\theta_1)},$$

which is negative if and only if $\theta_3 < \theta_1$. By induction, it is easy to see that

$$g_1^{(i)}(u) = c_i(1+u)^{\{i\theta_1-\theta_3-(i-1)\}/(1-\theta_1)}.$$

Hence, $\theta_3 < \theta_1$ and $\theta_1 > 1$ implies that $\{i\theta_1 - \theta_3 - (i-1)\}/(1-\theta_1) < 0$ for all $i \geqslant 1$ and thus $c_i > 0$ for odd i and $c_i < 0$ for even i. Arguing similarly regarding derivatives of $q_{21}\{p_{12}(u)\}$, the conditions of Theorem 1 are satisfied if and only if $\theta_3 < \theta_1$ and $\theta_3 < \theta_2$. In this case negative mass is assigned in some regions of space if $\theta_3 > \theta_1$ or $\theta_3 > \theta_2$, so that the condition is necessary. For our geopolitical example, the resulting constraint that between-household associations not exceed within-household associations is reasonable. However, such a constraint is probably not reasonable for parent–child clusters. To see this, redefine $\{1, 2\}$ as {mother, father} and $\{3, 4, 5\}$ as their children. Genetically, one expects that the parent–child association should exceed the father–mother association. For $I_{11} = \{1, 2\}$ and $I_{12} = \{3, 4, 5\}$, however, $I_{21} = \{1, 2, 3, 4, 5\}$ requires the parent–parent association to equal or exceed the parent–child association. Alternative level-1 partitions induce similar unreasonable constraints. While multivariate frailty models described in § 2 do accommodate this example, practitioners should note that such models roughly impose the constraint that between-subcluster associations be no larger than the associations within subclusters. Thus, in (3) the two parents should be assigned to separate, single-person subclusters.

5. Discussion

We have proposed a family of models for describing association in multivariate survival problems with multiple levels of clustering. Because we model association parametrically, our proposed family may be particularly appropriate when association is of key interest and insufficient data are obtainable to justify nonparametric estimation such as that described by Prentice & Cai (1992). A key advantage of our family is its specification subject only to existence of sufficiently regular Laplace transforms. This endows our formulation with a high degree of flexibility while retaining parametric power and interpretability. Moreover, it suggests the use of diagnostics to guide the choice of model in a given data-analytic situation (Oakes, 1989; Genest & Rivest, 1993).

We believe our models will be especially useful when researchers wish to impose a Cox (1972) or other regression structure on the marginal hazard functions corresponding to $S_j(t)$. Then, it would be straightforward in principle to write down a likelihood function for the observed data based on (5), accounting for censored observations in the usual way. It follows that one reasonable strategy for estimating parameters might be to maximise the likelihood in all of its arguments following Nielsen et al. (1992) and Murphy (1994). An alternative approach begins by estimating the marginal hazard parameters in a way which is robust to the structure of association (Lee, Wei & Amato, 1992; Liang, Self & Chang, 1993), and proceeds to estimate association parameters by the method of pseudo-maximum likelihood (Gong & Samaniego, 1981). Genest, Ghoudi & Rivest (1995) and Shih & Louis (1996) have recently discussed applying this approach to copula estimation, with nonparametric estimation of marginal survival functions. We are currently investigating the inferential properties of this approach in the regression setting.

One interesting goal for future research is to develop methods for estimating association parameters which do not require the full likelihood of the data, but retain the appealing measure of association provided by the conditional hazard ratio. In applications that we have encountered, bivariate association characteristics are often of greatest interest. This evokes thoughts of a methodology which requires only the specification of bivariate conditional hazard ratio structure, much as quasi-likelihood (Wedderburn, 1974) and generalised estimating equation (Liang & Zeger, 1986) approaches require only first- and second-order moments. Research into this issue is ongoing.

ACKNOWLEDGEMENT

This work was supported in part by a grant from the National Institutes of Health. We thank the editors and referees, whose comments greatly improved the presentation of the paper.

APPENDIX

Proofs

Proof of Proposition 1. Consider first the case of two levels. Properties (iii) and (iv) follow immediately from the monotonicity and continuity of Laplace transforms and their inverses under the assumptions. Since any Laplace transform maps 0 to 1 and converges to 0 as its argument converges to ∞, properties (i) and (ii) also hold. The general case of more than two levels follows immediately by induction.

Proof of Theorem 1. We need the following preliminary notation: let $f^{(k)}_{i_1 \ldots i_k}(v)$ denote the kth order partial derivative of f with respect to the variables $\{v_{i_1}, \ldots, v_{i_k}\}$; let '$\lfloor x \rfloor$' denote the integer part of x.

LEMMA A·1: *A chain rule for higher-order derivatives. Let $h(v)$ be of the form $h(v) = g\{f(v_1, \ldots, v_n)\}$.*

Then,

$$\frac{\partial^k}{\partial v_{i1}\dots\partial v_{ik}}h(v)=\sum_{i=1}^{k}g^{(i)}\{f(v)\}R_{k,i}(v)\quad(k=1,\dots,n),$$

with

$$R_{k,1}(v):=f^{(k)}_{i_1,\dots,i_k}(v),\quad R_{k,k}(v):=\prod_{u=1}^{k}f^{(1)}_{i_u}(v),$$

$$R_{k,i}(v):=\sum_{k_1=1}^{\lfloor k/i\rfloor}\sum_{k_2=k_1}^{(k-k_1)/(i-1)}\cdots\sum_{k_{i-1}=k_{i-2}}^{\lfloor(k-\sum_{u=1}^{i-2}k_u)/2\rfloor}\left\{\sum_c\prod_{u=1}^{i}p^{k_1\dots k_i}_{i_{u1},\dots,i_{uk_u}}f^{(k_u)}_{i_{u1},\dots,i_{uk_u}}(v)\right\}\quad(i=2,\dots,k)$$

for some set of constants $\{p^{k_1\dots k_i}_{i_{u1},\dots,i_{uk_u}}\}$, *each greater than* 0. *Above,* $\sum_c$ *denotes the sum over all possible combinations of* $\{(i_{11},\dots,i_{1k_1}),\dots,(i_{i1},\dots,i_{ik_i})\}$ *from* $(i_1,\dots,i_k)$, *and* $k_i=k-\sum_{u=1}^{i-1}k_u$.

The proof is a special case of Theorem 2.2.5 of Field (1976).

LEMMA A·2. *For each* $l=2,\dots,L$, $k=1,\dots,N_l$, $r\in M_{lk}$, *and for any* $A\subset I_{l-1,r}$,

$$\frac{\partial^{|A|}}{\prod_{j\in A}\partial t_j}q_{lk}\{S_{l-1,r}(t_{l-1,r})\}\begin{cases}\geqslant 0 & \text{for } |A| \text{ odd},\\ \leqslant 0 & \text{for } |A| \text{ even},\end{cases}$$

with $q_{lk}\{S_{l-1,r}(t_{l-1,r})\}$ *as defined in* (5).

Proof (by induction). For the case $l=2$, choose arbitrary $k, r, A\subset I_{1,r}$. If $|I_{1,r}|=1$, then $|A|=1$ and

$$\frac{\partial^{|A|}}{\prod_{j\in A}\partial t_j}q_{2k}\{S_{1,r}(t_{1,r})\}=\frac{\partial}{\partial t_j}q_{2k}\{S_j(t_j)\}=-q'_{2k}\{S_j(t_j)\}f_j(t_j)\geqslant 0.\tag{A1}$$

Suppose that $|I_{1,r}|>1$. Then

$$\begin{aligned}\frac{\partial^{|A|}}{\prod_{j\in A}\partial t_j}q_{2k}\{S_{1,r}(t_{1,r})\}&=\frac{\partial^{|A|}}{\prod_{j\in A}\partial t_j}q_{2k}\circ p_{1r}\left[\sum_{j'\in I_{1r}}q_{1r}\{S_{j'}(t_{j'})\}\right]\\&=(-1)^{|A|}(q_{2k}\circ p_{1r})^{(|A|)}\prod_{j\in A}q'_{1r}\{S_j(t_j)\}f_j(t_j).\end{aligned}$$

Because $q'_{1r}(v)$ is nonpositive on $v\in[0,1]$, the sign of $\partial^{|A|}q_{2k}\{S_{1r}(t_{1r})\}/\prod_{j\in A}\partial t_j$ is as claimed.

For the case $l>2$, suppose that the sign of $\partial^{|A|}q_{nk}\{S_{n-1,r}(t_{n-1,r})\}/\prod_{j\in A}\partial t_j$ is as claimed for each $A\subset I_{n-1,r}$, for all $r\in M_{nk}$, for all $k=1,\dots,N_n$, for each $n<l$. Choose arbitrary $k, r, A\subset I_{l-1,r}$. If $|I_{l-1,r}|=1$, then $\partial^{|A|}q_{lk}\{S_{l-1,r}(t_{l-1,r})\}/\prod_{j\in A}\partial t_j\geqslant 0$ exactly as in the case $l=2$. Supposing that $|I_{l-1,r}|>1$,

$$q_{lk}\{S_{l-1,r}(t_{l-1,r})\}=q_{lk}\circ p_{l-1,r}\left[\sum_{s\in M_{l-1,r}}q_{l-1,r}\{S_{l-2,s}(t_{l-2,s})\}\right].$$

Denote by a_s the number of indices in A belonging to $I_{l-2,s}$ and by 1 (.) the usual indicator function. Appealing to Lemma A·1 with

$$g:=q_{lk}\circ p_{l-1,r},\quad f:=\sum_{s\in M_{l-1,r}}q_{l-1,r}\{S_{l-2,s}(t_{l-2,s})\},$$

$\partial^{|A|}q_{lk}\{S_{l-1,r}(t_{l-1,r})\}/\prod_{j\in A}\partial t_j$ is as follows.

(i) $g^{(|A|)}\{f(v)\}R_{|A|,|A|}(v)$ *term.* By Theorem 1 conditions, $g^{(|A|)}\{f(v)\}$ has sign $(-1)^{|A|+1}$ on its nonzero range, with $v=t_{l-1,r}$. Further,

$$\frac{\partial f(t_{l-1,r})}{\partial t_j}=\frac{\partial}{\partial t_j}q_{l-1,r}\{S_{l-2,s}(t_{l-2,s})\}\geqslant 0$$

by the induction assumption. Hence, the sign of the $g^{(|A|)}\{f(v)\}R_{|A|,|A|}(v)$ term is as claimed.

(ii) $g^{(1)}\{f(v)\}R_{|A|,1}(v)$ *term*. Under the conditions of Theorem 1 $g^{(1)}\{f(v)\}>0$ on its nonzero range, with $v=t_{l-1,r}$. Further,

$$\frac{\partial^{|A|}}{\prod_{j\in A}\partial t_j}\sum_{s\in M_{l-1,r}} q_{l-1,r}\{S_{l-2,s}(t_{l-2,s})\}$$

has sign $(-1)^{|A|+1}$ if $\sum_{s\in M_{l-1,r}} 1(a_s>0)=1$, and equals 0 otherwise. Hence, the sign of the $g^{(1)}\{f(v)\}R_{|A|,1}(v)$ term is as claimed.

(iii) $g^{(i)}\{f(v)\}R_{|A|,i}(v)$ *terms for* $i=2,\ldots,|A|-1$. Under the conditions of Theorem 1 $g^{(i)}\{f(v)\}$ has sign $(-1)^{i+1}$ on its nonzero range with $v=t_{l-1,r}$. Using the '$R_{k,i}(v)$' term from Lemma A·1 with $k:=|A|$, notice that $f^{k_u}_{i_{u1},\ldots,i_{uk_u}}(t_{l-1,r})=0$ unless $\{i_1,\ldots,i_{uk_u}\}\subset I_{l-2,s}$ for some single $s\in M_{l-1,r}$. If

$$\{i_1,\ldots,i_{uk_u}\}\subset I_{l-2,s},\ s\in M_{l-1,r}, f^{(k_u)}_{i_{u1},\ldots,i_{uk_u}}(t_{l-1,r})$$

either equals 0 or has sign $(-1)^{k_u+1}$ by the induction condition. Then, $R_{|A|,i}(t_{l-1,r})$ either equals 0 or has sign $(-1)\sum_{u=1}^{i}(k_u+1)=(-1)^{|A|+i}$, so that the $g^{(i)}\{f(v)\}R_{|A|,i}(v)$ term either equals 0 or has sign $(-1)^{|A|+2i+1}$.

Thus, the sign of each term of $\sum_{i=1}^{k} g^{(i)}\{f(t_{l-1,r})\}R_{|A|,i}(t_{l-1,r})$ is consistent with the claim and the result of Lemma A·2 is proved.

Theorem 1 follows by applying Lemma A·1 with $g:=p_{L1}$ and $f:=\sum_{j=1}^{N_L} q_{Lj}\{S_{L-1,j}(t_{L-1,j})\}$. The conditions of the theorem and fact that $-p_{L1}$ satisfies those conditions then imply, according to Lemma A·2, that all odd-ordered mixed partial derivatives of $S(t)$ are nonpositive, and all even-ordered mixed partial derivatives are nonnegative, up to and including order m. □

REFERENCES

CHURCHILL, R. V. (1972). *Operational Mathematics*, 3rd ed. New York: McGraw-Hill.

CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–51.

COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–202.

FIELD, M. J. (1976). *Differential Calculus and its Applications*. New York: Van Nostrand Reinhold.

GENEST, C., GHOUDI, K. & RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–52.

GENEST, C. & MACKAY, R. J. (1986a). Copules archimédiennes of et familles de lois bidimensionnelles dont les marges sont données. *Can. J. Statist.* **14**, 145–59.

GENEST, C. & MACKAY, J. (1986b). The joy of copulas; bivariate distributions with uniform marginals. *Am. Statistician* **40**, 280–3.

GENEST, C. & RIVEST, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *J. Am. Statist. Assoc.* **88**, 1034–43.

GONG, G. & SAMANIEGO, F. J. (1981). Pseudo maximum likelihood. *Ann. Statist.* **9**, 861–9.

HUTCHINSON, T. P. (1981). Compound gamma bivariate distributions. *Metrika* **28**, 263–71.

JOE, H. (1993). Parametric families of multivariate distributions with given margins. *J. Mult. Anal.* **46**, 262–82.

LEE, E. W., WEI, L. J. & AMATO, D. A. (1992). Cox-type regression for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, Ed. J. P. Klein and P. K. Goel, pp. 237–47. Dordrecht: Kluwer Academic.

LIANG, K.-Y., SELF, S. G. & CHANG, Y.-C. (1993). Modelling marginal hazards in multivariate failure time data. *J. R. Statist. Soc.* B **55**, 441–53.

LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

MARSHALL, A. W. & OLKIN, I. (1988). Families of multivariate distributions. *J. Am. Statist. Assoc.* **83**, 834–41.

MURPHY, S. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. & SORENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–44.

OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

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

SHIH, J. H. & LOUIS, T. A. (1996). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* To appear.

VAUPEL, J. W., MANTON, K. G. & STALLARD, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography* **16**, 439–54.

WEDDERBURN, R. W. M. (1974). Quasilikelihood functions, generalized linear models and the Gauss–Newton method. *Biometrika* **61**, 439–47.

[*Received November* 1994. *Revised October* 1995]