*J. R. Statist. Soc.* A (1985)
**148,** *Part* 2, *pp.* 82–117

# Multivariate Generalizations of the Proportional Hazards Model

By DAVID CLAYTON† and JACK CUZICK

*Leicester University, UK* — *Imperial Cancer Research Fund Labs, London, UK*

[*Read before the* Royal Statistical Society *on Wednesday February 13th, 1985, at a meeting organized by the* Medical Section, *the President*, Dr W. F. Bodmer, *in the Chair*]

SUMMARY

A model for bivariate life-tables is considered with a single association parameter which is unaffected by monotone transformation of the marginal distributions. Methods for testing and estimating this parameter from right-censored sample pairs using only rank-order information are presented. The model is a generalization of the proportional hazards model and includes a random effect representing heterogeneity of "frailty" or proneness to failure. These methods are extended to allow for covariates and adapted to study the problem of intra-class association. The analysis of litter-matched and matched-pair failure-time data is discussed. Some uses of the methods in rank regression problems involving only one right-censored dependent variable are described and a test is proposed for proportional hazards against alternative error structures leading to converging hazards. Finally the methods are compared and validated by Monte-Carlo simulations.

*Keywords*: PROPORTIONAL HAZARDS; CENSORED DATA; BIVARIATE LIFE TABLE; NONPARAMETRIC ESTIMATION; MATCHED PAIRS; ASSOCIATION PARAMETER; INTRA-CLASS ASSOCIATION; SIBSHIP MODELS

## 1. INTRODUCTION

In the past decade there has been great interest and substantial advancement in the study of survival times for which only incomplete "censored" observations may be available. Particular attention has been given to developing non-parametric methods which permit comparison of survival curves in the presence of right censoring without the need to specify a detailed parametric form for the survival distributions. A concept fundamental to this approach is the proportional hazards model which was first applied in the two-sample problem (Mantel, 1966), and later was clearly formulated for the general regression problem by Cox (1972). Much of this early work was motivated by problems in the analysis of clinical trials and the development of methods for the two-sample, $k$-sample, and regression problems reflect these needs. The proportional hazards model is also of central importance in epidemiology, and is the concept underlying the "indirect" standardization of mortality and incidence rates, where a piecewise exponential model is implicitly assumed (Clayton, 1982). A similar concept is employed in the logistic regression analysis of case-control studies, but the basic hazard function is no longer available (Prentice and Breslow, 1978).

The success of these semi-parametric models in assessing treatment differences, prognostic factors and risk factors has stimulated an interest in the analysis of more detailed aspects of life histories. For example, one may wish to study the bivariate distribution of two time variables, or to include the time at which one event occurs in a regression analysis of some other life event.

† *Present address*: Dept of Community Health, Clinical Sciences Building, Leicester Royal Infirmary, P.O. Box 65, Leicester LE2 7LX, UK



0035–9238/85/14882 $2.00

Alternatively, the very occurrence of one life event might directly affect the hazard for another type of event.

A general framework for modelling life histories in terms of counting processes can be found in the work of Aalen *et al.* (1980). Our approach will be less general and will focus upon models for association in bivariate (and multivariate) survival distributions, and on the extension of regression models to allow for such association. We shall be concerned with association arising from heterogeneity of proneness, or "frailty", of individuals within a studied population rather than from that arising from direct effects of one life event upon another (the latter problem is discussed by Leurgans, Tsai, and Crowley, 1982). Finally we develop an extension of the proportional hazards regression model which incorporates both fixed and random terms.

Our motivation for studying these problems arose from analytical difficulties in clinical trials and epidemiological studies. For example, in cancer trials it is of interest to know if the interval from remission to relapse influences the subsequent interval from relapse to death. In heart disease epidemiology, the age of death from (or first attack of) heart disease in father–son pairs form a bivariate life distribution, with observations possibly doubly censored by, for example, mortality from other causes. Studies of breast cancer incidence in sisters provide the censored data analogue of the intra-class correlation problem. Again in cohort studies of breast cancer, two important covariates, viz. age at first pregnancy and age at menopause, will often be censored.

The next section describes a class of bivariate distributions characterized by an association parameter, but with arbitrary marginal distributions. The bivariate distribution is completely specified by this parameter and the two marginal distribution functions. We believe that this model is of interest even in the uncensored case and is applicable to bivariate problems other than those involving survival times. The next section reviews previous approaches to the problems of estimation and testing for this model, and the remaining sections deal with our current work. Section 4 develops the estimation theory for the case where there is no relationship between the marginal distributions of the two coordinates—we shall deal with the uncensored case in some detail and indicate the (minor) modifications necessary to deal with censoring. Section 5 extends this theory to allow for regression modelling of covariate effects. Finally, in Section 6 we examine what further information can be extracted by a non-parametric method when the two coordinates have the same (arbitrary) marginal distribution. This naturally generalizes to the case of several coordinates and allows the analysis of litter-matched experiments. A special case is the matched-pair survival data problem which has been the subject of some recent work, but which still awaits a fully satisfactory solution.

## 2. THE MODEL

We consider first the situation in which two time variables with unspecified marginal distributions are related by a single association parameter. Our goal is to construct the appropriate generalization of the classical proportional hazards model for this problem. As in the classical two-sample problem, the problem may be approached either as being concerned with the distribution of times, or as a contingency table problem concerned with the occurrence or non-occurrence of events.

We begin with the former approach. A natural model is to assume that each survival time is related to the same unobserved covariate by a proportional hazards model. If $(T_1, T_2)$ denotes the bivariate survival time with hazard function $(\lambda_1(t), \lambda_2(t))$ and $\omega$ denotes the unobserved covariate we assume that

$$\lambda_k(t \mid \omega) = \lambda_k^0(t) \exp(a_k\omega), \quad k = 1, 2,$$

where $\lambda_k(t \mid \omega)$ are the conditional hazard functions, $\lambda_k^0(t)$ are unknown baseline hazard rates and

$$\lambda_k(t) = E[\lambda_k(t \mid \omega)], \quad k = 1, 2,$$

are the expectations over $\omega$ of the conditional hazard function. Of course, complete specification of the model requires a distribution function for $\omega$.

This approach may also be formulated in terms of linear models and, as Prentice (1978) has pointed out for the univariate case, the monotone transformation

$$S_k = \log \Lambda_k^0(T_k)$$

applied to $(T_1, T_2)$ where

$$\Lambda_k^0(t) = \int_0^t \lambda_k^0(s)\, ds$$

yields the linear functional relationship

$$\begin{aligned} S_1 &= a_1 \omega + \epsilon_1 \\ S_2 &= a_2 \omega + \epsilon_2 \end{aligned} \tag{2.1}$$

where $(\epsilon_1, \epsilon_2)$ are independent random variables with (minus) extreme value distributions i.e.

$$\Pr(\epsilon > x) = \exp\{-\exp(x)\}.$$

Notice that each coordinate has been transformed by its own monotone transformation in this setup. Later we shall consider situations, such as matched pairs or litters, in which the coordinates are indistinguishable and this would be reflected by setting

$$\lambda_1^0(t) = \lambda_2^0(t) \text{ and } a_1 = a_2$$

so that the same transformation applied to both coordinates yields the linear model (2.1).

This linear model approach was used by Cuzick (1982) and by Wu (1982). The interpretation in terms of random shared relative risks was explored by Clayton (1978) who also investigated an approach based upon the limiting properties of contingency tables with a certain constant odds property. This was motivated by Mantel's treatment of the classical two-sample problem in which the proportional hazards model is the continuous time limit of a conditional logistic model in which the ratio of the odds of failure between groups remains constant over trials. Symbolically, if $P_{ij}$ denotes the relative frequency of failure at trial $i$ for subjects in group $j$ in this two sample problem, the model is

$$\theta = \frac{P_{i1} / \sum_{k>i} P_{k1}}{P_{i2} / \sum_{k>i} P_{k2}}, \quad \text{for all } i.$$

In the contingency table literature this model is known as the "continuation ratio" model (Fienberg, 1980; McCullagh and Nelder, 1983).

When we are interested in association between two discrete failure time variables then a natural analogue of this model suggests itself. If $P_{ij}$ now represents the relative frequency of failure of both components at trials $i$ and $j$ respectively, then the generalized continuation ratio model would require the constant odds ratio property

$$\theta = \frac{P_{ij} \left( \sum_{l>i} \sum_{k>j} P_{lk} \right)}{\left( \sum_{l>i} P_{lj} \right) \left( \sum_{k>j} P_{ik} \right)} \quad \text{for all } i, j.$$

The continuous time version of this model is obtained by grouping the time axes into discrete epochs and taking the limit as the durations of epochs approach zero. Assume $(T_1, T_2)$ have a joint density and let $F(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$ be the bivariate survivor function. Using the notation

$$F_1 = \partial F/\partial t_1, \qquad F_2 = \partial F/\partial t_2, \qquad F_{12} = \partial^2 F/\partial t_1 \partial t_2,$$

then the bivariate generalization of the proportional hazard model is

$$\theta = \frac{F\ F_{12}}{F_1 F_2}, \qquad \text{for all } t_1 \text{ and } t_2.$$

This has an interpretation in terms of conditional hazard functions, namely

$$\lambda_1(t_1 \mid T_2 = t_2) = \theta\ \lambda_1(t_1 \mid T_2 > t_2)$$

where $\lambda_1(t_1 \mid .)$ is the hazard function for $T_1$ conditional upon $T_2$; a similar relationship holds for the hazard of $T_2$ conditional upon $T_1$. A further representation of the model is in terms of what we shall call the *mortality potential*

$$H(t_1, t_2) = -\log F(t_1, t_2),$$

in the sense that its differentiation yields (conditional) hazards or "force of mortality" functions. For example, differentiation with respect to $t_1$ yields

$$H_1 = \lambda_1(t_1 \mid T_2 > t_2)$$

with a similar result for $H_2$. Written in terms of the mortality potential (which is simply a bivariate generalization of the integrated hazard or accumulated risk) our model becomes

$$H_{12} = -\gamma H_1 H_2, \quad \gamma \geqslant 0 \tag{2.2}$$

where $\gamma = \theta - 1$. For simplicity (and without loss of generality up to monotone transformations) we will take $T_1$ and $T_2$ to be non-negative random variables. The following theorem shows that (2.2) determines the joint distribution of $(T_1, T_2)$ up to monotone transformations of the marginal distributions.

*Theorem* 1. Assume $H(t_1, t_2)$ is twice continuously differentiable, $H(0, 0) = 0$ and let

$$\Lambda_1(t) = H(t, 0), \quad \Lambda_2(t) = H(0, t) \tag{2.3}$$

be the marginal cumulative hazard functions for $T_1$ and $T_2$ respectively. Then the unique solution of (2.2) subject to the boundary condition (2.3) is

$$\begin{aligned} H(t_1, t_2) &= \gamma^{-1} \log\{\exp(\gamma\Lambda_1(t_1)) + \exp(\gamma\Lambda_2(t_2)) - 1\}, & \gamma > 0 \\ &= \Lambda_1(t_1) + \Lambda_2(t_2), & \gamma = 0. \end{aligned}$$

The proof is given in Appendix 1.

As we are interested in nonparametric properties of the distribution we are free to choose the marginals at will. The choice $\Lambda_1(t) = \Lambda_2(t) = t$ yields unit exponential marginals and will be treated in detail below. We record that it has a bivariate survivor function

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

and joint density

$$\begin{aligned} F_{12}\ (t_1, t_2; \gamma) &\equiv \phi(t_1, t_2; \gamma) \\ &= \frac{(1+\gamma)\exp\{\gamma(t_1 + t_2)\}}{\{\exp(\gamma t_1) + \exp(\gamma t_2) - 1\}^{(2+1/\gamma)}} \end{aligned} \tag{2.4}$$

where $t_1 > 0, t_2 > 0$ and $\gamma > 0$.

This model gives independent coordinates when $\gamma = 0$ and maximally dependent coordinates as $\gamma \to \infty$. It does not permit negative association: i.e. negative values of $\gamma$ are not permitted.

Happily, the conflict between the two generalizations of the proportional hazards model has a satisfactory resolution. If the two error variables $(\epsilon_1, \epsilon_2)$ in (2.1) have (minus) extreme value distributions and it is assumed that the common covariate, $\omega$, has a (minus) log gamma distribution (i.e. $e^{-\omega}$ is a gamma variate with shape and scale parameter both equal to $\gamma^{-1}$ so that its mean is 1 and its variance is $\gamma$), and we set $a_1 = a_2 = 1$, then the bivariate distribution of $(T_1, T_2)$ in (2.1) agrees with that in (2.4) up to (separate) monotone transformation of each coordinate.

We remark in passing that, with the above assumptions concerning the random covariate $\omega$, $(T_1, T_2)$ in (2.1) have Pareto marginal distributions. This distribution is an example of a model for heterogeneity of hazard within a population. Such models have been studied by a number of authors including Vaupel *et al.* (1979) who introduced the term "frailties" for the multiplicative effects, $e^{-\omega}$. Hougaard (1984) gives some general results for heterogeneous frailty models and shows that the most tractable possibilities are gamma frailty models (to which class the models discussed here belong) and inverse-Gaussian frailty models. Hougaard also discusses the use of these models for the bivariate distribution of two latent failure times in the competing risk problem (Prentice *et al.*, 1978), a problem allied to that considered here.

We also remark that the model (2.1) may be generalized to include known covariates, $z_1$ and $z_2$ say, in addition to the random covariate $\omega$. Thus, conditional upon the unknown frailty, the effects of the fixed covariates may be represented in terms of the proportional hazard model. Unconditionally, however, the proportional hazards model no longer holds. In the two-sample problem, for example, the heterogeneity of frailty results in a decreasing ratio of marginal hazards with increasing survival time. Even in the univariate case, heterogeneous frailty models provide a useful alternative to the proportional hazards model; an application of these models to describe distributions of durations of unemployment is given by Lancaster and Nickell (1980). Further research along these lines is described in Heckman and Singer (1984) and Ridder and Verbakel (1983).

## 3. PREVIOUS APPROACHES TO INFERENCE

When the marginal cumulative hazard functions (2.3) are known parametric functions, $\Lambda_1(t; \alpha_1, \gamma)$ and $\Lambda_2(t; \alpha_2, \gamma)$ say, then the inference problem is straightforward, with three groups of parameters $\alpha_1$, $\alpha_2$ and $\gamma$. Clayton (1978) gave the log likelihood function for (doubly) censored data, and suggested bivariate extensions of the Weibull distribution and of the piecewise exponential distribution, again using gamma frailty assumptions. This derivation of the likelihood implicitly assumed type I censoring of each time variable and, although the same likelihood will certainly supply a sound basis for inference under less restrictive assumptions about the censoring mechanism, these conditions remain to be determined.

The specification of exponential or Weibull forms for the conditional distributions as suggested by Clayton (1978) leads to models in which the marginal distributions are Pareto or generalized Pareto distributions. The characterization described in the last section in which the marginal distributions are exponential was introduced by Oakes (1982), who discussed maximum likelihood estimation in the uncensored case. In this characterization the marginal integrated hazard functions are

$$\Lambda_1(t) = \alpha_1 t \qquad \Lambda_2(t) = \alpha_2 t$$

and the association parameter, $\gamma$, forms the third parameter of the model. Writing $(X_1^i, X_2^i)$ for the $i$th observation of $(T_1^i, T_2^i)$, we note that the score test for association is given by the first derivative of the log-likelihood with respect to $\gamma$ at $\gamma = 0$ i.e. by

$$\{\partial L/\partial\gamma\}_{\gamma=0} = \sum_{i=1}^{N} (\alpha_1 X_1^i - 1)(\alpha_2 X_2^i - 1)$$

which in this case, and more generally, may be written

$$\sum_{i=1}^{N} \{\Lambda_1(X_1^i) - 1\}\{\Lambda_2(X_2^i) - 1\}. \tag{3.1}$$

In the presence of censoring,

$$X_1^i = \min(T_1^i, C_1^i), \; X_2^i = \min(T_2^i, C_2^i),$$

where $(C_1^i, C_2^i)$ are potential censoring times. We must carry with $X_k^i$ censoring indicators $D_k^i$ taking the value 0 for a censored observation and 1 for an uncensored observation. At least for type I censoring, it may be easily shown that the score test for association with known marginal distributions is given by (3.1) with $-1$ replaced by $-D_k^i$, i.e. by

$$\sum_{i=1}^{N} \{\Lambda_1(X_1^i) - D_1^i\}\{\Lambda_2(X_2^i) - D_2^i\}. \tag{3.2}$$

Other work on this model has concentrated upon a semi-parametric approach in order to eliminate the need for parametric modelling of the marginal distributions. A natural approach is to base inference upon the marginal likelihood calculated from the joint probability of the two rank vectors. Oakes (1982) investigated this approach in the uncensored case, but was able to obtain an explicit expression for the marginal likelihood only for values of $N$ up to 3. Nevertheless, local approximations to the marginal likelihood have proved useful in deriving non-parametric tests for association. Cuzick (1982) and Wu (1982) investigated association tests for doubly censored data using the linear model (2.1). Working independently, both authors used a generalization of the marginal likelihood of ranks which was proposed by Kalbfleisch and Prentice (1973) and explored in relation to censored rank tests in the univariate case by Prentice (1978). We shall describe this likelihood formally in the next section, but here remark that it is a true marginal likelihood only for progressive type II censoring (Crowley, 1974). Cuzick and Wu showed that the form of the score test for no association ($\gamma = 0$) is determined by the distribution functions for the errors $(\epsilon_1, \epsilon_2)$, and is independent of the distribution of the unobserved covariate, $\omega$. For our model, the errors are (minus) extreme value variates and, in the uncensored case, the score test is given by (3.1) with $\Lambda_1(X_1^i)$ and $\Lambda_2(X_2^i)$ (which are transformations of $\{X_1^i\}$, $\{X_2^i\}$ to unit exponential variates) replaced by the expected order statistics from these (unit exponential) distributions corresponding to the ranks of the $\{X_1^i\}$ and $\{X_2^i\}$. The subtraction of 1 from these values corrects the scores to zero mean and yields what, following the work of Savage (1956), are known as "Savage scores".

In the censored case, this approach leads to a score test of the same form as (3.2), but with $\Lambda_k(X_k^i)$ replaced by corresponding estimates of the marginal cumulative hazards calculated from the censored observations by the method of Altshuler (1970) and Nelson (1969). Again, these values may be regarded as the expected exponential order statistics corresponding to the generalized ranks of the $X_k^i$, but under a progressive type II censoring scheme. The subtraction of $D_k^i$ corrects the scores to zero mean and yields what are known, following Peto and Peto (1972), as the "log-rank scores", or alternatively (as here) the "generalized Savage scores".

In the parametric setup, likelihood considerations lead to appropriate variance estimators for score statistics. Consistent estimates for the variance of censored rank tests for association have presented difficulties, however. For a random censorship model in which the censoring times $(C_1^i, C_2^i)$ are independent of each other and all other pairs, and their distribution does not depend on $i$, a permutational variance could be used. However, $C_1^i$ and $C_2^i$ are often dependent. Cuzick (1982) found an estimator which is valid for general error distributions and random censorship models. This estimate is complicated in general but reduces to a simple form for the proportional hazards model, viz., the variance of the score test based upon (3.2) can be consistently estimated by the number of doubly uncensored pairs. A similar result holds for the parametric bivariate exponential model (Oakes, 1982).

Clayton (1978) attempted to solve the semi-parametric problem in the Mantel-Cox spirit using a bivariate generalization of partial likelihood. This "likelihood" is the product of conditional probabilities evaluated over the cartesian product of the set of uncensored observations of $T_1$ and the set of uncensored observations of $T_2$. That such probabilities may be combined in this manner is suggested by standard results concerning the partitioning of the chi-squared test statistic for association in contingency tables (Lancaster, 1950; Plackett, 1981). However in this case the partitioning does not reflect a factorization of the underlying likelihood, and Oakes (1982) pointed out that the individual terms of Clayton's "likelihood" are not independent (or more generally, do not form a martingale difference sequence). Thus, although the test and estimation procedures developed by Clayton may retain some convenient properties, standard likelihood theory cannot be assumed and the variance estimates are optimistic.

Oakes proposed a third approach based upon Kendall's (1962) rank test for association. If we classify a pair $(i, j)$ as "concordant" if the signs of the differences $(X_1^i - X_1^j)$ and $(X_2^i - X_2^j)$ are the same, and "discordant" if they are different, then Kendall's $S$ is given by the total number of concordant pairs minus the total number of discordant pairs, thus having zero expectation under the null hypothesis. Oakes showed that, for our model, the ratio of the expectations of these totals is $\theta$, and that the ratio of the observed numbers of concordant to discordant pairs provides a consistent estimator for $\theta$. Oakes obtained an explicit expression for the variance of this estimator, but considered only the uncensored case. The generalization of his approach to doubly censored data would seem possible, simply omitting from the calculations those pairs in which the smaller of $X_1$ or $X_2$ is right censored, since such pairs cannot be classified as either concordant or discordant. This is in the spirit of Gehan's generalization of the Wilcoxon/Mann-Whitney test, but its consequences for the variances of the test statistic and estimator remain to be determined.

It is interesting to note that further generalization of Oakes' approach is possible. This involves assigning some weight, $W_{ij}$, to each pair and calculating, for the test statistic

$$U = \sum_{\substack{ij \\ \text{concordant}}} W_{ij} - \sum_{\substack{ij \\ \text{discordant}}} W_{ij} \tag{3.3}$$

and for the estimator,

$$\hat{\theta} = \sum_{\substack{ij \\ \text{concordant}}} W'_{ij} \Big/ \sum_{\substack{ij \\ \text{discordant}}} W'_{ij} \tag{3.4}$$

It may be shown that the test proposed by Clayton is of the form (3.3), with $W_{ij} = 1/N_{ij}$, and his maximum "likelihood" estimator is of the form (3.4) with $W'_{ij} = 1/(N_{ij} - 1 + \hat{\theta})$, where $N_{ij}$ is the total number of pairs with both individuals at risk at time

$$\{\min(X_1^i, X_1^j),\ \min(X_2^i, X_2^j)\}.$$

Clayton also proposed a locally efficient estimator in the spirit of the Mantel-Haenszel odds ratio estimator, with $W'_{ij} = W_{ij}$. Although the fallacy in the likelihood construction leads us to question the optimality of these statistics, nevertheless we might expect these weighted forms to be more efficient than the unweighted statistics proposed by Oakes.

## 4. ESTIMATION OF THE ASSOCIATION PARAMETER

It has been shown that an efficient test for association between two possibly censored time variables is based on generalized Savage scores. Asymptotically efficient estimation procedures can also be derived by replacing complicated multiple integrals in the MLE equations by scores. However fully efficient estimators do not use generalized Savage scores, but require scores which are adjusted away from them by an amount which depends on a crude estimate of the association parameter.

### 4.1. *A Likelihood*

To ease the notation we begin by studying the uncensored case. The modifications needed to accommodate censoring are similar to those for simpler well-understood situations and will be outlined below.

Assume we observe $N$ independent pairs $(X_1^i, X_2^i)$, $i = 1, \ldots, N$ which after a separate monotone transformation applied to each coordinate have the bivariate exponential density $\phi$ given by (2.4). When different transformations are applied to each coordinate, the maximal rank-invariant information about $\gamma$ is contained in the two rank vectors $R_1$ and $R_2$, where $R_1 = (R_1^1, \ldots, R_1^N)$ and $R_1^i$ is the rank of $X_1^i$ among all $(X_1^1, \ldots, X_1^N)$ and similarly for $R_2$. The marginal likelihood of ranks for $\gamma$ then takes the form

$$L(R_1, R_2; \gamma) = \int \cdots \int_{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}} \prod_{i=1}^{N} \phi(x_1^i, x_2^i; \gamma)\, dx_1^i\, dx_2^i, \tag{4.1}$$

where $\phi$ is given by (2.4) and $\{x_k \in R_k\}$, $k = 1, 2$, is taken to mean the set of $(x_k^1, \ldots, x_k^N)$ consitent with the observed rank vector $R_k$. This likelihood is of little computational value as it stands, but we will show that an asymptotically efficient estimate of $\gamma$ can be obtained by using the parametric likelihood based on the density (2.4) and replacing $(X_1^i, X_2^i)$ by appropriate scores.

*Theorem* 2. Let $l(x_1, x_2; \gamma) = \log \phi(x_1, x_2; \gamma)$. If $\hat{\gamma}$ maximizes $L(R_1, R_2; \gamma)$ and $\bar{\bar{\gamma}}$ solves (uniquely)

$$\sum_{i=1}^{N} l'(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \gamma) = 0 \tag{4.2}$$

where $l'$ denotes $\partial l/\partial\gamma$ and

$$\bar{\bar{x}}_k^i = E(X_k^i \mid R_1, R_2)$$
$$= \int \cdots \int_{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}} x_k^i \prod_{j=1}^{N} \phi(x_1^j, x_2^j; \gamma)\, dx_1^j\, dx_2^j \Bigg/ \int \cdots \int_{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}} \prod_{j=1}^{N} \phi(x_1^j, x_2^j; \gamma)\, dx_1^j dx_2^j \tag{4.3}$$

for $k = 1, 2, i = 1, \ldots, N$, then $\hat{\gamma} - \bar{\bar{\gamma}} = O(N^{-1} \log^2 N)$ so that $\bar{\bar{\gamma}}$ is asymptotically efficient. Here and below expectations conditional on $(R_1, R_2)$ also implicitly depend on $\gamma$.

This expression for $\bar{\bar{x}}_k^i$ is still intractable and also is dependent on $\gamma$. However we will give two computationally feasible approximations to $\bar{\bar{x}}_k^i$ below when $\gamma$ is known or estimated consistently. Then (4.2) and (4.3) can be used to provide an iterative estimate for $\gamma$. The procedure is first to set $\gamma = 0$ and compute $\bar{\bar{x}}_k^i$. In this case $\bar{\bar{x}}_k^i = \bar{x}_k^i \equiv E(X_k^i \mid R_k)$ does not depend on the ranks of the other coordinate. The scores $\bar{x}_k^i$ are the classical generalized Savage scores. These scores are used in (4.2) to produce a first approximate estimate $\bar{\gamma}$ of $\gamma$, which is then used in (4.3) to update the scores $\bar{\bar{x}}_k^i$, and produce a refined estimate $\bar{\bar{\gamma}}$. Further iteration is rarely necessary and in fact the estimate $\bar{\bar{\gamma}}$ performs better than all previous estimations in Monte Carlo simulations (see below). This is an approximate *EM* algorithm; only the ranks of the observations are known and the scale values are missing, and the algorithm fills out these incomplete data at each step of the iteration.

*Remark*. The result of Theorem 2 and the variance estimator (4.4) below are not restricted to the bivariate exponential distribution (2.4) but hold with minor modifications for quite general multivariate distributions, in particular they are valid for the bivariate Pareto form of the model (see 2.1).

*Proof of Theorem* 2. Let $L = L(R_1, R_2; \gamma)$ and using (4.1) expand the integrand contained in $\partial(\log L)/\partial\gamma$ about $(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i)$, $i = 1, \ldots, N$ to get

$$\frac{\partial(\log L)}{\partial\gamma} = \sum_{i=1}^{N} l'(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \gamma)$$

$$+\frac{1}{2} E\left\{ \sum_{i=1}^{N} \sum_{j=1}^{2} \sum_{k=1}^{2} (X_j^i - \bar{\bar{x}}_j^i)(X_k^i - \bar{\bar{x}}_k^i)(l'_{jk})^i \right\} \tag{4.4}$$

where $l' = \partial l/\partial\gamma$, $l'_{11} = \partial^2 l'/(\partial x_1)^2$ etc., and $(l'_{11})^i$ means that $l'_{11}$ is evaluated at some point between $(X_1^i, X_2^i)$ and $(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i)$. Now

$$|\, l'_{jk} \,| = O(X_1 + X_2), \text{ for } j, k = 1, 2$$

so that (see Appendix II)

$$|\, (l'_{jk})^i \,| = O(X_1^i + X_2^i + \bar{\bar{x}}_1^i + \bar{\bar{x}}_2^i), \quad i = 1, \ldots, N$$

and since, for $k = 1, 2$,

$$E(X_k^i)^2 = O(\bar{\bar{x}}_k^i)^2,$$

$$E(X_k^i - \bar{\bar{x}}_k^i)^4 = O\{E(X_k^i - \bar{\bar{x}}_k^i)^2\}^2$$

and

$$E(X_k^i - \bar{\bar{x}}_k^i)^2 = \operatorname{var}(X_k^i \mid R_1, R_2) \leqslant \operatorname{var}(X_k^i \mid R_k)$$

$$= E(X_k^i - \bar{x}_k^i)^2 = \sum_{j=1}^{I} (N-j+1)^{-2} = O(N-I+1)^{-1},$$

where $I$ is the rank of $X_k^i$ among the $(X_k^1, \ldots, X_k^N)$ and the last bound follows from the fact that when only the marginal ranks $R_k$ are given (or equivalently $\gamma = 0$), the $X_k^i$ can be viewed as exponential order statistics. (The expectations on the left are conditional on $(R_1, R_2; \gamma)$. These estimates can be used with the Cauchy-Schwartz inequality to bound the second term in (4.4) by a constant times

$$\sum_{i=1}^{N} (N-i+1)^{-1} \left( \sum_{j=1}^{i} (N-j+1)^{-1} \right) = O(\log^2 N).$$

It is easily checked that

$$\liminf_{N\to\infty} N^{-1} \sum_{i=1}^{N} l''(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \gamma) > 0$$

with probability one. Since $\partial \log L/\partial\gamma \mid_{\gamma=\hat{\gamma}}$ is zero, we can expand the first term in the right hand side of (4.4) about $\bar{\bar{\gamma}}$ and evaluate it at $\hat{\gamma}$ to obtain the desired result that $\bar{\bar{\gamma}} - \hat{\gamma}$ is $O_p(N^{-1} \log^2 N)$.

*Corollary*. If $\bar{\gamma}$ solves

$$\sum_{i=1}^{N} l'(\bar{x}_1^i, \bar{x}_2^i; \gamma) = 0,$$

then $\bar{\gamma} - \hat{\gamma}$ is $O_p(N^{-\frac{1}{2}})$, so that $\bar{\gamma}$ has positive asymptotic relative efficiency.

*Proof.* Expand $\Sigma\, l'(\bar{x}_1^i, \bar{x}_2^i; \gamma)$ about $(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}}), i = 1, \ldots, N$, to obtain

$$\sum_{i=1}^{N} \{(\bar{x}_1^i - \bar{\bar{x}}_1^i)\,(l_1')^i + (\bar{x}_2^i - \bar{\bar{x}}_2^i)\,(l_2')^i\} \Big/ \sum_{i=1}^{N} l''(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}}) \tag{4.5}$$

$$+ \text{ higher order terms,}$$

where $(l_k')^i$ is evaluated at $(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i, \bar{\bar{\gamma}})$. The estimates in the theorem can be used to show that $(\bar{x}_k^i - \bar{\bar{x}}_k) = O(N - I + 1)^{-\frac{1}{2}}$, $|(l_k')^i| = O(\bar{\bar{x}}_1^i + \bar{\bar{x}}_2^i)$ so that the numerator in (4.5) is $O(N^{\frac{1}{2}})$.

## 4.2. *A Variance Estimator*

We have shown that an efficient estimate for the non-parametric model can be obtained by inserting appropriate scores into the MLE equations. However the variance obtained by inserting these scores into the variance estimate based on the parametric Fisher information is too small except when $\gamma = 0$. Instead it is necessary to compute the observed Fisher information based on (4.1) and then insert scores (and their covariances) to obtain a consistent estimate. For $\gamma = \hat{\gamma}$ we have that

$$\frac{\partial^2 (\log L)}{\partial \gamma^2} = \frac{L''(\hat{\gamma})}{L(\hat{\gamma})}$$

$$= \underset{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}}{\int \cdots \int} [\{\sum_i l''(x_1^i, x_2^i; \hat{\gamma})\} + \{\sum_i l'(x_1^i, x_2^i; \hat{\gamma})\}^2]\; d\phi(\hat{\gamma}) \Big/ \underset{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}}{\int \cdots \int} d\Phi(\hat{\gamma}), \tag{4.6}$$

where

$$d\Phi(\gamma) = \prod_{i=1}^{N} \phi(x_1^i, x_2^i; \gamma)\, dx_1^i\, dx_2^i. \tag{4.7}$$

The first term in (4.6) can be adequately approximated by $\Sigma l''(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \hat{\gamma})$ and the second term would be negligible if the $(x_1^i, x_2^i)$ were replaced by $(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i)$. However a Taylor expansion around these values and based at $\gamma$ gives a non-negligible term

$$E\left[\left\{\sum_{i=1}^{N} (x_1^i - \bar{\bar{x}}_1^i)\, l_1'\, (\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}}) + (x_2^i - \bar{\bar{x}}_2^i)\, l_2'\, (\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}})\right\}^2 \Bigg| R_1, R_2\right] = b^{\mathrm{T}} \Sigma\, b$$

where $\Sigma$ is the conditional covariance matrix of $(x_1^1, \ldots, x_1^N, x_2^1, \ldots, x_2^N)$ given the rank vectors $R_1$ and $R_2$,

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

has components

$$b_k^i = l_k'\, (\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}}) \qquad i = 1, \ldots, N; k = 1, 2$$

and $l_k'$ is the partial derivative of $l'$ with respect to $x_k$.

A computationally feasible approximation for $\Sigma$ which gives an asymptotically equivalent quadratic form is given in Appendix III. Thus a variance estimator for $\bar{\bar{\gamma}}$ (or $\hat{\gamma}$) is

$$\left[ \sum_{i=1}^{N} l''(\bar{\bar{x}}_1^i, \bar{\bar{x}}_2^i; \bar{\bar{\gamma}}) + b^T \Sigma b \right]^{-1}.$$

Since $b^T \Sigma b$ is positive it has the effect of reducing the Fisher information and increasing the variance. Since $l_1'(x_1, x_2; 0) = l_2'(x_1, x_2; 0) = 0$, this term will be negligible when $\gamma = 0$, as it must to agree with the variance estimate for *testing* $\gamma = 0$. However it will make a non-trivial contribution when $\gamma \neq 0$.

### 4.3. *Modifications for Censored Data*

For censored data the observation times $(X_1^i, X_2^i)$ (censored or uncensored) must be augmented by the indicator functions $(D_1^i, D_2^i)$ for uncensored observations. The rank vectors $R_1$, $R_2$ must now be interpreted as generalized rank vectors (Prentice, 1978) so that $R_k^i = (J_k^i, D_k^i)$ where $J_k^i$ denotes the number of *uncensored* $X_k^m$ less than or equal to $X_k^i$, and $D_k^i$ is the indicator for $X_k^i$ being uncensored. Other important items are the risk sets defined as $\mathcal{R}_k^i = \{j: X_k^j \geqslant X_k^i\}$ when $X_k^i$ is uncensored. When $X_k^i$ is censored its risk set coincides with that of the largest uncensored observation less than or equal to $X_k^i$. We denote their cardinalties by $N_k^i = |\mathcal{R}_k^i|$. When there is no censoring $N_k^i = N + 1 - R_k^i$. The bivariate density in the likelihood (4.1) requires generalization to accommodate singly and doubly censored pairs. Define

$$\phi(x_1, x_2, d_1, d_2; \gamma) = \begin{array}{ll} \phi(x_1, x_2; \gamma) & , d_1 = d_2 = 1, \\ -F_1(x_1, x_2; \gamma) & , d_1 = 1, d_2 = 0, \\ -F_2(x_1, x_2; \gamma) & , d_1 = 0, d_2 = 1, \\ F(x_1, x_2; \gamma) & , d_1 = d_2 = 0, \end{array}$$

where $(d_1, d_2)$ are censoring indicators corresponding to $(x_1, x_2)$ and $F, F_1$ etc. are the bivariate survivor function of Section 2 and its derivatives; explicit expressions are given in Appendix II. If we omit explicit mention of $(d_1, d_2)$ in the notation for $\phi$ and regard $(R_1, R_2)$ as generalized rank vectors, then the likelihood (4.1) can be reinterpreted as a likelihood for a Type II progressive censoring model. In this case the integral should be interpreted as only being carried out over the uncensored variables, and times for censored variables should be interpreted as referring to the largest smaller uncensored time. Further details and discussion of this likelihood can be found in Cuzick (1982). It can also be shown that under general conditions (4.1) can be viewed as an approximate likelihood for a random censorship model.

The estimation procedure is unchanged for censored data. The generalized Savage (logrank) scores are given by the well known expression

$$\bar{x}_k^i = \sum_{X_k^j \leqslant X_k^i} (N_k^j)^{-1}$$

and

$$\bar{\bar{x}}_k^i - \bar{x}_k^i = \underset{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}}{\int \dots \int} (x_k^i - \bar{x}_k^i)\, d\Phi(\gamma) \Big/ \underset{\substack{\{x_1 \in R_1\} \\ \{x_2 \in R_2\}}}{\int \dots \int} d\Phi(\gamma), \tag{4.8}$$

where $d\Phi(\gamma)$ is given by (4.7). Note that $\bar{x}_k^i$ and $\bar{\bar{x}}_k^i$ estimate the time of the largest *uncensored* observation less than or equal to $x_k^i$ (or zero if there is no such value).

### 4.4. *The Pareto Form of the Model*

Alternatively we may work with a Pareto form of the model which is based more closely upon (2.1). For this representation, the likelihood can be written as

$$L(R_1, R_2; \gamma) = \int_{\xi} \underset{\substack{\{y_1 \in R_1\} \\ \{y_2 \in R_2\}}}{\int \cdots \int} \prod_{i=1}^{N} f(\xi^i y_1^i) f(\xi^i y_2^i) (\xi^i)^{(D_1^i + D_2^i)} \, dy_1^i \, dy_2^i$$

$$\times \prod_{i=1}^{N} g(\xi^i; \gamma) \; d\xi^i \tag{4.9}$$

where $f$ is the unit exponential density function for uncensored observations or survivor function for censored observations (they coincide, as it happens) and the usual interpretation of $y^i$ and $dy^i$ prevails for censored values. The function $g$ denotes the gamma density with parameters $(\gamma^{-1}, \gamma^{-1})$. The $\xi^i$ are the "frailties" and are related to the $\omega^i$ of (2.1) by $\xi^i = \exp(-\omega^i)$. The change from $x$'s to $y$'s emphasizes that, in this form of the model, $y$ has Pareto rather than exponential marginals. Transformation to unit exponential marginals can be carried out by

$$x = \{\log(1 + \gamma y)\}/\gamma, \tag{4.10}$$

the integrated hazard function for the Pareto distribution.

We note that the inner part of (4.9) is identical to Cox's partial likelihood for known $\{\xi^i\}$. Alternatively, (4.9) may be integrated with respect to $\{\xi^i\}$ to yield

$$L(R_1, R_2; \gamma) = \underset{\substack{\{y_1 \in R_1\} \\ \{y_2 \in R_2\}}}{\int \cdots \int} \prod_{i=1}^{N} \phi(y_1^i, y_2^i, D_1^i, D_2^i, \gamma) \, dy_1^i \, dy_2^i, \tag{4.11}$$

where the factors $\phi$ are defined as in Section 4.3 (their form is given in Appendix II). Again integration is carried out only over uncensored variables.

The estimator $\bar{\gamma}$ introduced in Section 4.1 may be obtained by maximizing the parametric likelihood which forms the inner part of (4.11) using scores $(\bar{y}_1^i, \bar{y}_2^i)$ which are obtained from the generalized Savage scores $(\bar{x}_1^i, \bar{x}_2^i)$ using the transformation inverse to (4.10).

### 4.5. *Computation of the Bar-Bar Scores*

We have investigated several approximate methods for computing the scores $\bar{\bar{x}}_k^i$ or $\bar{\bar{y}}_k^i$. We have used two approaches and each one may be applied in either form of the model, although we shall introduce our notation for the exponential form. Both approaches also lead to approximations to the conditional variance-covariance matrix, $\Sigma$, used in the variance estimate of Section 4.2; these are given in Appendix III.

We start by introducing the important function

$$\psi(x_1, x_2, d_1, d_2; \gamma) = \log \frac{\phi(x_1, x_2, d_1, d_2; \gamma)}{\phi(x_1, x_2, d_1, d_2; 0)}$$

and, making use of the change of measure formula

$$d\Phi(\gamma) = \left(\frac{d\Phi(\gamma)}{d\Phi(0)}\right) d\Phi(0),$$

we may write the scores

$$\bar{x}_k^i = \frac{E\left\{x_k^i \exp\left(\sum_{j=1}^{N} \psi^j\right) \middle| R_1, R_2, \gamma = 0\right\}}{E\left\{\exp\left(\sum_{j=1}^{N} \psi^j\right) \middle| R_1, R_2, \gamma = 0\right\}} \tag{4.12}$$

where $\psi^j = \psi(x_1^j, x_2^j, d_1^j, d_2^j; \gamma)$ (Appendix II). Notice that although the numerator and denominator of (4.11) are expectations of complicated variables, the probabilistic structure is simple ($\gamma = 0$) and they may be approximated by Taylor expansion of

$$\sum_{j=1}^{N} \psi^j. \tag{4.13}$$

Such expansions are in terms of the derivatives of $\{\psi\}$ with respect to $x_1^j, x_2^j, j = 1, \ldots, N$ and we shall use a subscript notation so that $\psi_1$, $\psi_2$ represent (vectors of) first derivatives and $\psi_{11}$, $\psi_{12}$ and $\psi_{22}$ represent (matrices of) second derivatives.

Our first approach was to expand (4.13) to second order around $(\bar{x}_1^j, \bar{x}_2^j)$ and to make a normal approximation for the distributions of the spacings of the order statistics under the simple probabilistic structure ($\gamma = 0$). This leads to an expression for the $\bar{x}_k^i$ which contains the inverse of a $2N \times 2N$ matrix. Alternatively, the $\bar{x}_k^i$ can be viewed as the solution of $2N$ linear simultaneous equations which may be solved iteratively although convergence is not guaranteed. Written in the iterative form, these equations are

$$\bar{\bar{x}}_k^i = \sum_{X_k^j \leqslant X_k^i} \frac{D_k^j}{(N_k^j)^2} \left( \sum_{l \in \mathcal{R}_k^j} (1 + \tilde{\psi}_k^l) \right) \tag{4.14}$$

where

$$\begin{pmatrix} \tilde{\psi}_1 \\ \tilde{\psi}_2 \end{pmatrix} = \begin{pmatrix} \bar{\psi}_1 \\ \bar{\psi}_1 \end{pmatrix} + \begin{pmatrix} \bar{\psi}_{11} & \bar{\psi}_{12} \\ \bar{\psi}_{21} & \bar{\psi}_{22} \end{pmatrix} \begin{pmatrix} \bar{\bar{x}}_1 - \bar{x}_1 \\ \bar{\bar{x}}_2 - \bar{x}_2 \end{pmatrix} \tag{4.15}$$

which may be regarded as approximations to the first derivatives $\psi_k$ evaluated at $(\bar{\bar{x}}_1, \bar{\bar{x}}_2)$.

Our second approach involves expansion of (4.13) to first order only, but around $(\bar{\bar{x}}_1, \bar{\bar{x}}_2)$ rather than $(\bar{x}_1, \bar{x}_2)$. Heuristically, we might expect this to represent the same order of approximation as our first approach. Now there is no need of the normal approximation, but of necessity this leads to an iterative method which seems to be numerically more stable than that implied by (4.14) and (4.15) although it seems less amenable to theoretical development. Some algebra shows this approach to lead to the (non-linear) equations

$$\bar{\bar{x}}_k^i = \sum_{X_k^j \leqslant X_k^i} D_k^j \left\{ \sum_{l \in \mathcal{R}_k^j} (1 - \bar{\psi}_k^l) \right\}^{-1} \tag{4.16}$$

which may be solved iteratively starting from $(\bar{x}_1, \bar{x}_2)$. Comparison of this with (4.14) and (4.15) shows that the two methods are numerically similar when the discrepancies $(\bar{\bar{x}}_k^i - \bar{x}_k^i)$ are small.

The second method is also suggested by other more heuristic arguments. The first of these views the transformation of the observations $X_k^i$ to the scores $\bar{\bar{x}}_k^i$ (whose marginal distributions are known) as being piecewise linear with discontinuities at the observed failure times. This argument parallels that of Breslow (1974) in deriving a nonparametric estimator of the baseline hazard function of the proportional hazards model. A similar approach has been used by Bennett (1983a, 1983b) for the univariate rank regression problem with logistic errors.

However, perhaps the most compelling heuristic argument arises with the Pareto form of the model. Here it may be shown (Appendix II) that

$$1-\psi(y_1,y_2,d_1,d_2;\gamma)=\frac{1+\gamma(d_1+d_2)}{1+\gamma(y_1+y_2)}$$

and some algebra shows that when $(y_1^i, y_2^i, D_1^i, D_2^i)$ are known, the $\xi^i$ are independent gamma variables with shape parameter $(D_1^i+D_2^i+\gamma^{-1})$ and scale $(y_1^i+y_2^i+\gamma^{-1})$, so that

$$E(\xi^i \mid R_1,R_2)=E\left(\frac{1+\gamma(D_1^i+D_2^i)}{1+\gamma(y_1^i+y_2^i)}\,\middle|\,R_1,R_2\right)$$

and (4.16) becomes

$$\bar{\bar{y}}_k^i=\sum_{Y_k^j\leqslant Y_k^i} D_k^j\left(\sum_{l\in \mathcal{R}_k^j}\bar{\bar{\xi}}_k^l\right)^{-1}$$

where

$$\bar{\bar{\xi}}_k^l=\frac{1+\gamma(D_1^l+D_2^l)}{1+\gamma(\bar{\bar{y}}_1^l+\bar{\bar{y}}_2^l)} \tag{4.17}$$

may be thought of an "imputed frailties".

This leads to an alternative conceptual view of our proposed method. Since the only dependence of the likelihood (4.9) upon $\gamma$ is in the function $f$, it is easily shown that

$$\frac{\partial(\log L)}{\partial\gamma}=\sum_{i=1}^{N} E\left\{\frac{\partial \log f(\xi^i;\gamma)}{\partial\gamma}\mid R_1,R_2)\right\}. \tag{4.18}$$

Also

$$\frac{\partial \log f(\xi;\gamma)}{\partial\gamma}=\frac{1}{\gamma^2}\{\xi-\log\xi-1+\log\gamma+\Psi(\gamma^{-1})\} \tag{4.19}$$

where $\Psi(\eta)$ is the digamma (psi) function $\Gamma'(\eta)/\Gamma(\eta)$, and this scoring method is formally equivalent to approximation of (4.18) by

$$\frac{\partial(\log L)}{\partial\gamma}=\sum_{i=1}^{N} E\left\{\frac{\partial \log f(\xi^i;\gamma)}{\partial\gamma}\mid R_1,R_2)\right\}_{\xi^i=\bar{\bar{\xi}}^i}.$$

A variance estimator can be computed as before. Using (4.9), a little algebra shows that

$$\frac{\partial^2\log L}{\partial\gamma^2}=E\left\{\sum_{i=1}^{N}\frac{\partial^2}{\partial\gamma^2}\log f(\xi^i;\gamma)+\left(\sum_{i=1}^{N}\frac{\partial}{\partial\gamma}\log f(\xi^i;\gamma)\right)^2\right\} \tag{4.20}$$

where the expectations are conditional on $R_1$ and $R_2$.

The first term in (4.20) can be approximated by replacing $\xi^i$ by $\bar{\bar{\xi}}^i$ and, after expansion about the $\bar{\bar{\xi}}^i$, the second term has the non-trivial approximation

$$E\left\{\sum_{i=1}^{N}(\xi^i-\bar{\bar{\xi}}^i)\frac{\partial^2}{\partial\gamma\partial\xi}\log f(\xi;\gamma)\Big|_{\xi=\bar{\bar{\xi}}^i,\,\gamma=\bar{\bar{\gamma}}}\right\}^2$$

which requires the evaluation of the covariances of the $\xi^i$ given $R_1$ and $R_2$. This can be done by conditioning on the $(y_1^i, y_2^i)$. To first order, one can treat the $\xi^i$ as being conditionally orthogonal with variances given by

$$\gamma E\left\{\frac{1+\gamma(D_1^i+D_2^i)}{\{1+\gamma(y_1^i+y_2^i)\}^2}\,\middle|\, R_1, R_2\right\} + \operatorname{var}\left\{\frac{1+\gamma(D_1^i+D_2^i)}{1+\gamma(y_1^i+y_2^i)}\,\middle|\, R_1, R_2\right\}$$
$$\cong \bar{\bar{\gamma}}\,\frac{1+\bar{\bar{\gamma}}(D_1^i+D_2^i)}{\{1+\bar{\bar{\gamma}}(y_1^i+y_2^i)\}^2}\,.$$

## 5. FIXED COVARIATES

The development in the preceding sections can be extended to handle observed covariates. The linear model (2.1) now becomes

$$S_1 = \beta_1^{\mathrm{T}} z_1 + \omega + \epsilon_1 \qquad S_2 = \beta_2^{\mathrm{T}} z_2 + \omega + \epsilon_2 \tag{5.1}$$

where $(z_1, z_2)$ are observed covariate vectors which may be different for the two coordinates, $(\beta_1, \beta_2)$ are vectors of regression coefficients whose components are assumed to be different in general, even for the same covariate and, as before, $\omega$ is a shared random effect independent of $(\epsilon_1, \epsilon_2)$ which have independent standard extreme value error distributions. A minor modification would allow some of the regression coefficients to be the same in both coordinates. Another model with the same covariates in each coordinate and identical regression coefficients will be developed in the next section.

Because of rank invariance, it is not possible to estimate constant (hazard level) effects in either coordinate. Estimation of $(\beta_1, \beta_2, \gamma)$ is based on a modification of the scoring technique used previously. The following fact is basic to this scoring procedure:

*Theorem* 3. Let $T_i$ be independent exponential random variables with parameters $\lambda_i$, $i = 1, \ldots, N$. Define the random index $I$ by $T_I = \min_i(T_i)$. Then $I$ and $T_I$ are independent and the distribution of $T_I$ is exponential with parameter $\Sigma\lambda_i$ so that

$$E\{T_I \mid I\} = \left(\sum_{i=1}^{N} \lambda_i\right)^{-1}.$$

From this it follows that if $Y_k^i = \exp(S_k^i)$ is uncensored and $\Delta_k^i$ denotes the difference between $Y_k^i$ and the largest uncensored $Y_k^j$ strictly less than $Y_k^i$ (zero if no such value), then conditional on $(R_1, R_2, \omega)$, the $\Delta_k^i$ are independent exponential random variables with means

$$E(\Delta_k^i \mid R_1, R_2, \omega) = E(\Delta_k^i \mid R_k, \omega)$$
$$= \left[\sum_{j \in \mathcal{R}_k^i} \exp\{-(\omega^j + \beta_k^{\mathrm{T}} z_k^j)\}\right]^{-1}.$$

(Strictly speaking, we must also assume that all censoring takes place instantaneously after the previous uncensored value, as in Type II censoring. However for realistic censoring mechanisms this only affects the expectation to order $N^{-2}$ and can be ignored.)

A number of important new results can be derived for the simpler problem of only one dependent variable and these are considered first.

### 5.1. *One Dependent Variable*

In this case we omit subscripts from all variables and write

$$S = \beta^{\mathrm{T}} z + \omega + \epsilon, \tag{5.2}$$

where now $(\omega + \epsilon)$ can be treated as the log of a Pareto error variable. By fixing $\gamma$, we have a model for rank regression for any member of this error family. The proportional hazard regression model corresponds to $\gamma = 0$. Testing procedures for $\beta = 0$ in a general linear rank regression model have been presented by Prentice (1978). A one parameter sub-family of these weighted logrank tests has been studied by Harrington and Fleming (1982). These tests are precisely the score tests for the partial likelihood based on (5.2) and our model provides an interpretation of their tests in terms of heterogeneous $\gamma$-frailties. In addition our approach can also be used for multivariate point and interval estimation. The choice $\gamma = 1$ corresponds to a logistic error distribution and is worthy of special note.

It is interesting to note that the proportional hazards model is in some sense an extreme point and that the Pareto and log-logistic error structures can be thought of as arising from the addition of further unmeasured population heterogeneity.

By not fixing $\gamma$ in advance the amount of population heterogeneity can be estimated. In particular, the score test for $\gamma = 0$ provides a simple useful test for proportional hazards against the alternative of converging hazards (which is a consequence of population heterogeneity). Because the proportional hazards model is extremal within our family, no test for diverging hazards is available. However, tests for other Pareto (in particular log-logistic) error structures can also be developed as well as tests for $\beta = 0$ at some estimated $\gamma$.

We begin with the form of the likelihood used in Section 4.4:

$$L(R; \beta, \gamma) = \int_{\xi} \int \cdots \int_{\{y \in R\}} \prod_{i=1}^{N} \phi(y^i \xi^i \theta^i)\, (\xi^i \theta^i)^{D^i} f(\xi^i; \gamma)\, dy^i\, d\xi^i$$

where $\xi^i$, $\phi$ and $f$ are as in (4.9) and $\theta^i$ are the relative risk functions $\exp(-\beta^{\mathrm{T}} Z^i)$. Then

$$\partial \log L/\partial\gamma = \sum_{i=1}^{N} E\left\{ \frac{\partial \log f(\xi^i; \gamma)}{\partial \gamma} \mid R \right\} \tag{5.3}$$

$$\partial \log L/\partial\beta = \sum_{i=1}^{N} E\{Z^i(-D^i + \xi^i y^i \theta^i) \mid R\}. \tag{5.4}$$

When $\beta$ is fixed, it is easily verified that, given the vectors $y$ and $R$, the $\xi^i$ are conditionally independent gamma variables with shape parameters $\gamma^{-1} + D^i$ and scale $\gamma^{-1} + y^i\theta^i$. It follows that

$$\begin{aligned} E(\xi^i y^i \mid R) &= E\{y^i E(\xi^i \mid y, R) \mid R\} \\ &= E\left\{ \frac{y^i(1+\gamma D^i)}{1+\gamma y^i \theta^i} \,\middle|\, R \right\}. \end{aligned}$$

Thus, as at (4.17), we can rewrite (5.3) and (5.4) in terms of $y^i$ alone as

$$\frac{\partial \log L}{\partial \gamma} = \frac{1}{\gamma^2} \left[ \sum_{i=1}^{N} E\left\{ \frac{1+\gamma D^i}{1+D^i y^i \theta^i} + \log(1+\gamma D^i y^i \theta^i) \,\middle|\, R \right\} - \sum_{i=1}^{N} (1+\gamma D^i) \right] \tag{5.5}$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{N} Z^i\, E\left\{ -D^i + \frac{(1+\gamma D^i)\, y^i \theta^i}{1+\gamma y^i \theta^i} \,\middle|\, R \right\}. \tag{5.6}$$

To find $(\bar{\bar{\beta}}, \bar{\bar{\gamma}})$ we now set these equations equal to zero, replace $y^i$ by an estimate of $\bar{\bar{y}}^i = E(y^i \mid R)$ and solve iteratively: set $\gamma = 0$, compute $\bar{\bar{\beta}}$ from (5.6) as in a proportional hazards model, compute $\bar{\bar{y}}$ as shown below, compute $\bar{\bar{\gamma}}$ from (5.5) with the current estimates of $\beta$ and $\bar{\bar{y}}$ and iterate as necessary.

As with method 2 of the last section, it can be shown that

$$E(y^i \mid \xi, R) = \sum_{Y^j \leqslant Y^i} \frac{D^j}{\sum_{l \in \mathcal{R}^j} \xi^l \theta^l}$$

$$E(\xi^i \mid y, R) = \frac{1 + \gamma D^i}{1 + \gamma y^i \theta^i}$$

which suggests the iterative approximation

$$\bar{\bar{y}}^i \cong \sum_{Y^j \leqslant Y^i} \frac{D^j}{\sum_{l \in \mathcal{R}^j} \bar{\bar{\xi}}^l \theta^l}$$

$$\bar{\bar{\xi}}^i \cong \frac{1 + \gamma D^i}{1 + \gamma \bar{\bar{y}}^i \theta^i}\,.$$

When $\gamma = 0$, the Pareto and exponential forms of the model coincide and the scores can be computed exactly (when $\beta$ is known). Not surprisingly in the light of the discussion of the previous section, they turn out to be identical to Breslow's (1974) estimator of the integrated baseline hazard function,

$$E(y^i \mid R, \theta, \gamma = 0) = \sum_{Y^j \leqslant Y^i} \frac{D^j}{\sum_{l \in \mathcal{R}^j} \theta^l}\,. \tag{5.7}$$

When $\gamma \neq 0$, we may define $\bar{y}$ scores as we did in Section 4.4, i.e. as those obtained from (5.7) by a transformation inverse to (4.10).

The information matrix for $L$ is given by

$$\frac{\partial^2 \log L}{\partial \gamma^2} = E\left\{ \sum_{i=1}^{N} \frac{\partial^2 \log f}{\partial \gamma^2} + \left( \sum_{i=1}^{N} \frac{\partial \log f}{\partial \gamma} \right)^2 \middle| R \right\} - \left( \frac{\partial \log L}{\partial \gamma} \right)^2$$

$$\frac{\partial^2 \log L}{\partial \gamma \partial \beta} = E\left\{ \left( \sum_{i=1}^{N} \frac{\partial \log f}{\partial \gamma} \right) \left( \sum_{i=1}^{N} Z^i (-D^i + \xi^i y^i \theta_i) \middle| R \right) \right\} - \left( \frac{\partial \log L}{\partial \gamma} \right) \left( \frac{\partial \log L}{\partial \beta} \right)$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta^{\mathrm{T}}} = E\left[ -\left\{ \sum_{i=1}^{N} \xi^i y^i \theta^i (Z^i)^{\otimes 2} \right\} + \left\{ \sum_{i=1}^{N} (-D^i + \xi^i y^i \theta^i) Z^i \right\}^{\otimes 2} \middle| R \right] - \left( \frac{\partial \log L}{\partial \beta} \right) \left( \frac{\partial \log L}{\partial \beta^{\mathrm{T}}} \right)^{\mathrm{T}}$$

where $Z^{\otimes 2}$ denotes $ZZ^{\mathrm{T}}$, $\partial \log f/\partial\gamma$ is given by (4.17) and

$$\frac{\partial^2 \log f}{\partial\gamma^2} = \frac{1}{\gamma^3}\{3-\gamma^{-1}\,\Psi'(\gamma^{-1})-2(\xi-\log\xi+\log\gamma+\Psi(\gamma^{-1}))\}.$$

As before the last term in each of these expressions is asymptotically negligible. The last term in each expectation can be evaluated by expanding about bar-bar scores, and the terms involving second derivatives can be evaluated by replacing all $\xi^i$ by the appropriate function of $y^i$ (by making use of their conditional gamma distribution), and then evaluating the $y^i$ at the $\bar{\bar{y}}^i$ score values. Alternatively one can begin with (5.6) and (5.7), which have no $\xi^i$ terms and take further derivatives of these expressions. Both approaches are cumbersome in general, but simplification occurs in some cases.

In particular a test for proportional hazards ($\gamma = 0$) at some estimated $\hat{\beta}$ can be based on the $\bar{y}$ scores given at (5.7). From (5.5) a little calculation shows that the test statistic is

$$T = \sum_{i=1}^{N} E\{-D^i y^i\hat{\theta}^i + (y^i\hat{\theta}^i)^2/2 \mid R\}$$

$$\cong \frac{1}{2}\left\{\sum_{i=1}^{N} (\bar{y}^i\hat{\theta}^i - D^i)^2 - (D^i)\right\} \tag{5.8}$$

where $\hat{\theta}^i = e^{-\hat{\beta}^{\mathrm{T}} z^i}$ and $\hat{\beta}$ is estimated at $\gamma = 0$ by the usual proportional hazards methods (Cox, 1972). This test has mean approximately zero when $\gamma = 0$ and is of the same form as a test for overdispersion when the null hypothesis requires independent unit exponential variables (see Cox and Lewis, 1966). The power of the test will depend on the amount of variation induced in $Y$ by $\beta^{\mathrm{T}}Z$. When $\gamma = 0$ the estimators for $\gamma$ and $\beta$ are asymptotically independent so a variance estimate for (5.8) can be based on $-\partial^2 \log L/\partial\gamma^2$ and takes the form

$$\operatorname{var}(T) \cong \sum_{i=1}^{N} (\bar{y}^i\hat{\theta}^i)^2\,[2\bar{y}^i\hat{\theta}^i/3 - D^i]$$

*Remarks.*

(1) Taking $\gamma = 0$ gives the proportional hazards model. In this case our method is an exact *EM*-algorithm and provides an alternative computational scheme for maximum "partial" likelihood estimation. This method is particularly convenient for implementation via *GLIM* (see Clayton and Cuzick, 1985, for details).

(2) Fitting the fully parametric regression model with Pareto errors may easily be achieved in the computer program *GLIM* (Baker and Nelder, 1978) using an extension of the method described by Aitkin and Clayton (1980) for exponential errors. The indicator variables, $D^i$, are declared as $y$-variate, with binomial error structure, the number of trials being $(D^i + \gamma^{-1})$, and the logarithms of the observed times are declared as offset. The replacement of times by bar-bar scores would appear fairly straightforward as a macro so that the computations described in this section would appear to be conveniently carried out in *GLIM*. This will be reported in detail elsewhere.

(3) In the case of fixed $\gamma$, our estimation procedure differs from that proposed by Pettitt (1982) in that he would expand the logarithm of the likelihood about $\beta = 0$ to quadratic terms and use the approximate log-likelihood for estimation. Such a procedure leads to biased estimates in general and does not enjoy the asymptotic properties of our method. The $R$-estimation procedure discussed by Huber (1977) provides another general approach, but would require considerable further development to be tractable in the present context. Also, it requires more than basic rank information for estimation.

### 5.2. *Two Dependent Variables*

We return now to the model (5.1). The interpretation of $\gamma$ in this model differs from that in (5.2) in that now $\gamma$ is a measure of association between $T_1$ and $T_2$ after adjustment for $Z_1$ and $Z_2$, whereas in (5.1) $\gamma$ was a measure of population heterogeneity. The extent to which the interpretation of $\gamma$ as an association parameter is confounded by population heterogeneity which is *not* shared by the two coordinates (i.e. independent variables $\omega_1$ and $\omega_2$ added to $S_1$ and $S_2$ respectively) is in need of further study.

The likelihood for the ranks of $N$ independent observations from (5.1) is given by

$$L(R_1, R_2; \beta_1, \beta_2, \gamma) =$$

$$\int_{\xi} \int_{\{y_1 \in R_1\}} \cdots \int_{\{y_2 \in R_2\}} \prod_{i=1}^{N} \phi(\xi^i y_1^i \theta_1^i)\, \phi(\xi^i y_2^i \theta_2^i)\, (\xi^i \theta_1^i)^{D_1^i}\, (\xi^i \theta_2^i)^{D_2^i}\, f(\xi^i; \gamma)\, dy_1^i dy_2^i d\xi^i,$$

where $\theta_k^i$ are the relative risk functions $\exp(-\beta_k^{\mathrm{T}} Z_k^i)$ and with the usual conventions for censored observations. The *MLE* equations are obtained from

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^{N} E\left\{ \frac{\partial \log f(\xi^i; \gamma)}{\partial \gamma} \,\middle|\, R_1, R_2 \right\}$$

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^{N} E\{Z_k^i (-D_k^i + \xi^i y_k^i \theta_k^i) \mid R_1, R_2\}, \quad k = 1, 2,$$

with the appropriate modifications if some of the $\beta_1$ are constrained to equal some of the $\beta_2$. The estimation procedure is similar to that developed previously, the main difference being that the $\xi^i$ conditional on $y_1^i, y_2^i, R_1, R_2$ have independent gamma distributions with shape and scale parameters

$$(D_1^i + D_2^i + \gamma^{-1}), \quad (y_1^i \theta_1^i + y_2^i \theta_2^i + \gamma^{-1})$$

so that

$$\bar{\bar{\xi}}^i \equiv E(\xi^i \mid R_1, R_2) = E\left( \frac{1 + \gamma(D_1^i + D_2^i)}{1 + \gamma(y_1^i \theta_1^i + y_2^i \theta_2^i)} \,\middle|\, R_1, R_2 \right).$$

Scoring is similar to the previous subsection. When the $\xi^i$ are known or taken as fixed, estimation of $\beta_1, \beta_2$ can be carried out by the usual proportional hazards algorithms. In particular the score test for $\gamma = 0$ (no association) takes the form

$$T = \sum_{i=1}^{N} (\bar{y}_1^i \hat{\theta}_1^i - D_1^i)(\bar{y}_2^i \hat{\theta}_2^i - D_2^i)$$

$$+ \sum_{i=1}^{N} \{(\bar{y}_1^i \hat{\theta}_1^i - D_1^i)^2 - D_1^i\} + \sum_{i=1}^{N} \{(\bar{y}_2^i \hat{\theta}_2^i - D_2^i)^2 - D_2^i\},$$

$$\operatorname{var}(T) \cong \sum_{i=1}^{N} \{D_1^i D_2^i - (D_1^i + D_2^i)(\bar{y}_1^i \hat{\theta}_1^i + \bar{y}_2^i \hat{\theta}_2^i)^2/2 + (\bar{y}_1^i \hat{\theta}_1^i + \bar{y}_2^i \hat{\theta}_2^i)^3/3\},$$

where the $\hat{\theta}_k^i$ are as at (5.8). This test has the form of an inner product between the $(\bar{y}^i \hat{\theta}^i - D^i)$

scores and two terms which measure population heterogeneity in the marginals. In view of the possible confounding of association with unassociated marginal heterogeneity a more robust test, when this is suspected, might be simply the inner product

$$\sum_{i=1}^{N} (\bar{y}_1^i \hat{\theta}_1^i - D_1^i)(\bar{y}_2^i \hat{\theta}_2^i - D_2^i).$$

Another possibility is to estimate $\gamma_1$ and $\gamma_2$ for the marginals separately as in Section 5.1, score each marginal according to the estimates of $\beta_k$ and $\gamma_k$, $k = 1, 2$ and then correlate these scores.

## 6. MATCHED PAIRS AND SIBLING PROBLEMS

In some circumstances, such as the matched pairs setup of Holt and Prentice (1974) or Woolsen and Lachenbrach (1980), the randomized block experiments of Sampford and Taylor (1959), or the litter matched experiments analysed by Mantel, Bohidar and Ciminera (1977), and Mantel and Ciminera (1979), members of the sibships (matched groups) are indistinguishable save (possibly) for the values of covariates. In such circumstances only a single universally applied monotone transformation will be necessary to reduce all variables to a canonical scale. This implies that the interleaved rankings of the combined data set contains more information than the separate marginal rank vectors. Also, any covariates should have the same regression coefficients for all individuals. We shall use the word sibship to denote matched group more generally, and use superscripts to denote sibship for consistency with earlier sections. For consistency, subscript should denote individual *within* sibship, but for greater flexibility we shall use subscripts to denote individual *regardless* of sibship. The model can now be written as

$$S_i = \beta^{\mathrm{T}} z_i + \omega^j + \epsilon_i \tag{6.1}$$

when the $i$th individual belongs to the $j$th sibship, denoted $i \in \mathbf{S}^j$. Note that the vector $\beta$ is independent of position in the sibship. As before we take the inter-sibship variation $\omega$ to be log-gamma and the intra-sibship (individual) variation $\epsilon$ to be extreme value. The data consist of the covariate values, a combined generalized rank vector for survival times denoted $R$, and a vector specifying sibship membership, denoted $\mathbf{S}$. The likelihood takes the form

$$L(R, \mathbf{S}; \beta, \gamma) = \int \cdots \int_{\{y \in R\}} \int_{\xi} \prod_{j} \prod_{i \in \mathbf{S}^j} \phi(\xi^j y_i e^{\beta z_i}) f(\xi^j; \gamma)\, d\xi^j\, dy_i$$

where $j$ indexes sibships, $\xi^j = \exp(-\omega^j)$ and the usual conventions for censored values apply. *MLE* equations, variance estimates and approximate scores can be developed as before, except that the expectations are now conditional on the combined ranking $R$.

Two important cases are (a) a measure of association with no covariates and (b) randomized treatment allocation with matched pairs. In the first case the likelihood can be written as

$$L(R; \gamma) = \int \cdots \int_{\{y \in R\}} \prod_{j} [\int (\xi^j)^{D_\cdot^j} e^{-y_\cdot^j} f(\xi^j; \gamma)\, d\xi^j]\, dy,$$

where

$$dy = dy_1 \ldots dy_N, \quad D_\cdot^j = \sum_{i \in \mathbf{S}^j} D_i, \quad y_\cdot^j = \sum_{i \in \mathbf{S}^j} y_i\ .$$

This leads to the *MLE* equation

$$\frac{\partial \log L}{\partial \gamma} = \sum_j E\left\{ \frac{\partial \log f(\xi^j; \gamma)}{\partial \gamma} \,\middle|\, R \right\} = 0$$

which can be rewritten in terms of the $y_i$ by noting that, conditional on all the $y_i$, $i \in \mathcal{R}^j$, $\xi^j$ is a gamma variable with parameters $(\gamma^{-1} + D_\cdot^j)$ and $(\gamma^{-1} + y_\cdot^j)$. Thus

$$\frac{\partial \log L}{\partial \gamma} = \gamma^{-2} \sum_j \left[ E\left\{ \frac{1+\gamma D_\cdot^j}{1+\gamma y_\cdot^j} + \log(1+\gamma y_\cdot^j) \,\middle|\, R \right\} - \{1 + \Psi(D_\cdot^j + \gamma^{-1}) - \Psi(\gamma^{-1})\} \right]. \tag{6.2}$$

Point estimation can be carried out by setting (6.2) equal to zero and replacing the $y_i$ by the scores $\bar{y}_i$ and solving iteratively as before. The approximate scores test for $\gamma = 0$ is easily obtained since now the $\bar{\bar{y}}_i = \bar{y}_i$ are simply the logrank scores for the combined sample. This leads to the test statistic

$$T = \frac{1}{2}\left\{ \sum_j (\bar{y}_\cdot^j - D_\cdot^j)^2 - \sum_j D_\cdot^j \right\} \tag{6.3}$$

which has an interpretation as the difference between observed and expected intra-class variation.

Testing and interval estimation require a variance estimate which can be based on an approximation to

$$-\frac{\partial^2 \log L}{\partial \gamma^2} = E\left\{ \sum_i \frac{\partial^2 \log f(\xi^i)}{\partial \gamma^2} + \left( \sum_i \frac{\partial \log f(\xi^i)}{\partial \gamma} \right)^2 \,\middle|\, R \right\} + \left( \frac{\partial \log L}{\partial \gamma} \right)^2$$

$$\cong \sum_i -\frac{\partial^2 \log f(\bar{\xi}^i)}{\partial \gamma^2} + b^{\mathrm{T}} \Sigma b,$$

where $b = (b_1, \ldots, b_N)$ and

$$b_i = \left. \frac{\partial^2 \log f}{\partial \gamma \partial \xi^i} \right|_{y_i = \bar{y}_i}$$

and $\Sigma$ is the conditional covariance matrix of the $y_i$ given $R$ which can be computed as before. In particular when $\gamma = 0$, $b = 0$ so that a variance estimator for the test (6.3) is

$$\operatorname{var}(T) \cong \frac{1}{3} \sum_j \{2(\bar{y}_\cdot^j)^3 - 3D_\cdot^j(\bar{y}_\cdot^j)^2 + D_\cdot^j(D_\cdot^j - 1/2)(D_\cdot^j - 1)\}.$$

Another important special case is randomized treatment allocation in matched pairs. In this case we can use equations (6.1) and interpret $z_i$ as an indicator variable for treatment. Then $\beta$ is a scalar which measures relative survival differences associated with treatment and $\gamma$ accounts for variation shared within pairs. Additional covariates are easily accommodated. Estimation follows the usual prescription. In particular the test for $\beta = 0$ with an estimated $\gamma$ takes the form

$$T = \sum_{j=1}^{N} \bar{\xi}^j U^j - \sum_{i=1}^{2N} Z_i D_i,$$

where $j$ indexes the $N$ pairs, $i$ indexes the $2N$ individuals,

$$\bar{\xi}^j = \frac{1 + \bar{\bar{\gamma}} D_\cdot^j}{1 + \bar{\bar{\gamma}} \bar{y}_\cdot^j}, \qquad U^j = \sum_{i \in S^j} (Z_i \bar{y}_i),$$

and all scores are obtained by setting $\beta = 0$ and using the univariate $MLE$ estimate $\bar{\bar{\gamma}}$ for $\gamma$. A variance estimate takes the form

$$\left. \sum_{j=1}^{N} \bar{\xi}^j U^j - \left( \frac{\partial \log L}{\partial \beta} \right)^2 \right|_{\beta = 0, \gamma = \bar{\bar{\gamma}}}.$$

## 7. NUMERICAL RESULTS

In this section we investigate the behaviour of the methods of Section 4 by Monte Carlo simulations. Pairs of variables, $(X_1, X_2)$, distributed according to the bivariate exponential distribution with association parameter $\gamma$ were generated from pairs of independent uniform pseudo-random numbers, $(U_1, U_2)$, by the transformations

$$x_1 = -\log(u_1)$$

$$x_2 = \frac{1}{\gamma} \log \left\{ 1 + \frac{(1-p)}{p} (u_1)^{-\gamma} \right\},$$

where

$$p = (u_2)^{\gamma/(1+\gamma)}.$$

We generated 1000 datasets for each of four assumptions for sample size. The first three types of dataset consisted of $N = 50$, 100 or 200 pairs of observations, with the largest 10 values respectively of each variable censored, and the final type had $N = 100$ with the largest 50 censored. We took $\gamma = 1$, corresponding to the moderate degree of association of heart disease incidence/ mortality in fathers and sons (Clayton, 1978) and of breast cancer incidence in mothers and daughters (unpublished work). This corresponds with an exponential distribution of frailties. We estimated $\gamma$ for each dataset using five methods described in this paper: (a) the fully parametric method, taking the marginal means as unknown; (b) the method proposed in Section 4, using the bar (generalized Savage) scores; (c) the method of Section 4 using bar-bar scores calculated by our second approach of Section 4.5; (d) and (e) the concordance estimator (3.4) using, respectively, Oakes' (unit) weights and Clayton's non-iterative weights. In further simulations, our alternative approach to the computation of the bar-bar scores gave almost identical results, but were more prone to convergence problems. Other simulations using the Pareto "imputed frailty" scores also gave results almost indistinguishable from those obtained under (c). The concordance estimators required more computation and no results are presented for the largest sample size. Iterative refinement of the weights for Clayton's form of this estimator has a negligible effect upon the estimates for these sample sizes and was not attempted.

Table 1 shows the results of these simulations. All the non-parametric estimators performed well, particularly under the heaviest censoring where all achieved full efficiency relative to the parametric method. Unexpectedly, in the remaining cases, the estimator $\gamma$ outperformed the other non-parametric estimators.

It would appear that, when the parametric *ML* estimate is low, then $\bar{\gamma}$ tends to be even lower. The refinement of the bar scores to bar-bar scores seems to correct this, but at the cost of an increase in errors of overestimation. We conclude that, for these values of $N$ and $\gamma$, the methods we have investigated for computation of the bar-bar scores are not sufficiently accurate to capture the small amount of information theoretically available. The computationally simpler method using bar scores would, therefore, be recommended for practical purposes.

The concordance estimators also performed surprisingly well. As expected, the weighted (Clayton) estimator was rather more efficient than the unweighted (Oakes) form.

## 8. CONCLUDING REMARKS

Starting from the problem of nonparametric estimation of the association parameter of a bivariate generalization of the proportional hazards model, we have been led to consider a wider class of models. This class is an extension of the proportional hazards regression model of Cox (1972) which allows a random effect (distributed as the log of a gamma variate) in addition to the fixed effects of covariates. At its simplest, this allows rank regression analysis in a generalized Pareto family including both proportional hazards and proportional odds (logistic) models. More complex is the bivariate model, which allows separate regression models for both variables and associated errors. The sibship problem is one example of a multivariate model. The most complex

TABLE 1

*Distribution of estimates from 1000 simulations, gamma = 1*

| Method | Mean | Median | S.D. | I.Q. range |
|---|---|---|---|---|
| *N = 50, 10 censored* | | | | |
| (a) Parametric | 1.063 | 1.041 | 0.403 | 0.763–1.328 |
| (b) Bar scores | 1.063 | 1.027 | 0.423 | 0.754–1.322 |
| (c) Bar-bar scores | 1.092 | 1.053 | 0.438 | 0.776–1.358 |
| (d) Oakes | 1.050 | 1.020 | 0.454 | 0.730–1.295 |
| (e) Clayton | 1.063 | 1.021 | 0.429 | 0.754–1.324 |
| *N = 100, 10 censored* | | | | |
| (a) Parametric | 1.033 | 1.011 | 0.254 | 0.855–1.186 |
| (b) Bar scores | 1.028 | 1.003 | 0.268 | 0.844–1.196 |
| (c) Bar-bar scores | 1.046 | 1.019 | 0.272 | 0.859–1.213 |
| (d) Oakes | 1.028 | 1.012 | 0.285 | 0.831–1.203 |
| (e) Clayton | 1.034 | 1.006 | 0.275 | 0.847–1.207 |
| *N = 200, 10 censored* | | | | |
| (a) Parametric | 1.005 | 0.993 | 0.171 | 0.880–1.105 |
| (b) Bar scores | 1.007 | 1.000 | 0.184 | 0.877–1.125 |
| (c) Bar-bar scores | 1.018 | 1.011 | 0.185 | 0.886–1.136 |
| *N = 100, 50 censored* | | | | |
| (a) Parametric | 1.044 | 1.007 | 0.459 | 0.709–1.327 |
| (b) Bar scores | 1.043 | 1.011 | 0.458 | 0.711–1.329 |
| (c) Bar-bar scores | 1.045 | 1.011 | 0.459 | 0.713–1.331 |
| (d) Oakes | 1.036 | 1.011 | 0.458 | 0.708–1.309 |
| (e) Clayton | 1.039 | 1.009 | 0.456 | 0.713–1.323 |

model would permit general multivariate gamma frailty distributions, normalized so that each marginal frailty has shape parameter equal to its scale parameter. This extension will be developed elsewhere.

Our approach is fully nonparametric in the dependent variable(s) and consists of, firstly, transforming the observed responses so as to have known marginal distributions rank order scores and, secondly, estimating the parameters of the model by maximum likelihood treating the scores as if they were genuine observations with known marginal distributions. As the heterogeneity of frailty in the population (i.e. the variance of the random effect) approaches zero, the efficient scores approach generalized Savage (logrank) scores, but the computation of efficient scores when frailty is heterogeneous presents difficulties. We have suggested two possible approaches, but there is a need for further work both on the properties of these methods and upon possible improvements.

The field of application of such models is wider than the failure time problems which motivated this research. The possibility of fully nonparametric testing and estimation of treatment effects in split plot designs is of particular interest. An example in medical statistics is the analysis of the two-period cross-over clinical trial which is complicated by the occurrence of both between- and within-patient errors. This problem is commonly avoided by relating some contrast between the first and second period responses to the treatment order, but a rank invariant analysis of this type can only use sign contrasts and must be expected to be inefficient.

There is considerable scope for extension of the methods described in this paper. Clayton (1978) discussed inference from case-control studies of familial aggregation of disease using the model of Section 2. Although his likelihood is inaccurate, our simulations suggest that it may be applicable in this important special case. A study of the present approach under the extremely heavy right censoring occurring in epidemiological incidence studies, and of its implications for the analysis of case-control studies would be of some interest and importance.

## REFERENCES


Aalen, O. O., Borgen, O., Keiding, N. and Thormann, J. (1980) Interaction between life history events. Non-parametric analysis for prospective and retrospective data in the presence of censoring. *Scand. J. Statist.*, **7**, 161–171.

Aitkin, M. and Clayton, D. G. (1980) The fitting of exponential Weibull and extreme value distributions to complex censored survival data using GLIM. *Appl. Statist.*, **29**, 56–63.

Altshuler, B. (1970) Theory for the measurement of competing risks in animal experiments. *Math. Biosc.*, **6**, 1–11.

Baker, R. J. and Nelder, J. A. (1978) *The GLIM System, Release 3*. Oxford: Numerical Algorithms Group.

Bennett, S. (1983a) Log-logistic regression models for survival data. *Appl. Statist.*, **32**, 165–171.

——— (1983b) Analysis of survival data by the proportional odds model. *Stats. in Medicine*, **2**, 273–277.

Breslow, N. E. (1974) Covariance analysis of censored survival data. *Biometrics*, **30**, 89–99.

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.

——— (1982) The analysis of prospective studies of disease aetiology. *Communications in Statistics*, **11**, 2129–2155.

Clayton, D. and Cuzick, J. (1985) The *EM* algorithm for Cox's regression model using GLIM. *Appl. Statist.*, **34**, in press.

Cox, D. R. (1972) Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B, **34**, 187–220.

Cox, D. R. and Lewis, P. A. (1966) *The Statistical Analysis of a Series of Events*. London: Methuen.

Crowley, J. (1974) A note on some recent likelihoods leading to the log rank test. *Biometrika*, **61**, 957–961.

Cuzick, J. (1982) Rank tests for association with right-censored data. *Biometrika*, **69**, 351–364.

Fienberg, S. E. (1977) *The Analysis of Cross-Classified Categorical Data*. Cambridge Mass.: MIT Press.

Harrington, D. P. and Fleming, T. R. (1982) A class of rank test procedures for censored survival data. *Biometrika*, **69**, 553–566.

Heckman, J. and Singer, B. (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, **52**, 271–319.

Holt, J. D. and Prentice, R. L. (1974) Survival analysis in twin studies and matched pair experiments. *Biometrika*, **61**, 17–30.

Hougaard, P. (1984) Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, **71**, 75–83.

Huber, P. J. (1977) *Robust Statistical Procedures*. Regional Conference Series *27*, SIAM, Philadelphia.

Kalbfleisch, J. D. and Prentice, R. L. (1973) Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 267–278.

Kendall, M. G. (1962) *Rank Correlation Methods*, Third Edition. London: Griffin.

Lancaster, H. O. (1950) The exact partition of chi-squared and its application to the problem of pooling of small expectations. *Biometrika*, **37**, 267–278.

Lancaster, T. and Nickell, S. (1980) The analysis of re-employment probabilities for the unemployed (with Discussion). *J. R. Statist. Soc.* A, **143**, 141–165.

Leurgans, S., Tsai, W-Y., and Crowley, J. (1982) Freund's bivariate exponential distribution and censoring. In *Survival Analysis* (J. Crowley and R. A. Johnson, eds.), 230–242. Institute of Mathematical Statistics lecture notes monograph series Volume 2.

Mantel, N. (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.*, **50**, 163–170.

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

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

McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models*. London: Methuen.

Nelson, W. B. (1969) Hazard Plotting for incomplete failure data. *J. Qual. Techn.*, **1**, 27–52.

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

Peto, R. and Peto, J. (1972) Asymptotically efficient rank invariant test procedures (with Discussion). *J. R. Statist. Soc.* A, **135**, 185–206.

Pettitt, A. N. (1982) Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Soc.* B, **44**, 234–243.

Plackett, R. L. (1981) *The Analysis of Categorical Data*. London: Griffin.

Prentice, R. L. (1978) Linear rank tests with right censored data. *Biometrika*, **65**, 167–179.
Prentice, R. L. and Breslow, N. E. (1978) Retrospective studies and failure time models. *Biometrika*, **65**, 153–158.
Prentice, R. L., Kelbfleisch, J. D., Petersen, A. V., Fluornoy, N., Farewell, V. T. and Breslow, N. E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.
Ridder, G. and Verbakel, W. (1983) On the estimation of the proportional hazard model in the presence of unobserved heterogeneity, Report AE 22/83, Dept. of Econometrics, University of Amsterdam.
Sampford, M. R. and Taylor, J. (1959) Censored observations in randomized block experiments. *J. R. Statist. Soc.* B, **21**, 214–237.
Savage, I. R. (1956) Contributions to the theory of rank order statistics—the two sample case. *Ann. Math. Statist.*, **27**, 590–615.
Vaupel, J. W., Manton, K. G. and Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439–454.
Woolson, R. F. and Lachenbruch, P. A. (1980) Rank tests for censored matched pairs. *Biometrika*, **67**, 597–606.
Wu, S.-C. (1982) Rank tests for independence based upon partially right censored pairs. *Communications in Statistics*, **11**, 2207–2216.

## APPENDIX I
### *Proof of Theorem 1*

The proof is trivial when $\gamma = 0$. Take $\gamma > 0$. First assume that $\Lambda_1'$ and $\Lambda_2'$ are strictly positive on $(0, \infty)$. Then $H_1 > 0$ and $H_2 > 0$ on $R_2^+$ for if $H_2(t_1, t_2) = 0$ for some $(t_1, t_2)$, then $H_{12}(t_1, t_2) = 0$ and (2.2) implies $H_2(t, t_2) = 0$ for all $t \geqslant 0$. In particular $H_2(0, t_2) = 0$ in contradiction to our assumption. Thus we can write (2.2) as

$$H_1 = -\gamma^{-1} H_{12}/H_2 = -\gamma^{-1} \frac{\partial}{\partial t_1} (\log H_2).$$

Integrating this with respect to $t_1$ gives

$$H = -\gamma^{-1} \log H_2 + G(t_2) \tag{I.1}$$

for some function $G$. Evaluating this at the boundary $(0, t_2)$ gives

$$\Lambda_2(t_2) = H(0, t_2) = -\gamma^{-1} \log \{\Lambda_2'(t_2)\} + G(t_2)$$

and substituting this into (I.1), we have the first order ordinary differential equation

$$H_2 = A(t_2) \exp(-\gamma H) \tag{I.2}$$

with initial condition $H(t_1, 0) = \Lambda_1(t_1)$ where

$$A(t_2) = \Lambda_2'(t_2) \exp\{\gamma \Lambda_2(t_2)\}. \tag{I.3}$$

The substitution $z = \exp(-\gamma H)$ in (I.2) leads to the equation

$$z' = -\gamma A(t) z^2$$

which is easily solved to yield

$$\exp(\gamma H) = z = \{\gamma \int A(s)\, ds + \text{const}\}^{-1}.$$

Using (I.3), we find

$$\int A(s)\, ds = \gamma^{-1} \exp\{\gamma \Lambda_2(t_2)\} + \text{const},$$

so that

$$H(t_1, t_2) = \gamma^{-1} \log [\exp\{\gamma \Lambda_2(t_2)\} + F(t_1)]$$

for some function $F$. The fact that $F(t_1) = \exp\{\gamma \Lambda_1(t_1)\} - 1$ follows from the boundary condition.

In general, when $\Lambda_1$ and $\Lambda_2$ are assumed only to have continuous derivatives let

$$x = \Lambda_1(t_1), \quad y = \Lambda_2(t_2)$$

so that

$$H_1 = H_x \Lambda_1'$$
$$H_2 = H_y \Lambda_2'$$
$$H_{12} = H_x H_y \Lambda_1' \Lambda_2'$$

and (2.2) is equivalent to

$$H_{xy} = -\gamma H_x H_y \quad \text{or } \Lambda_1' = 0 \text{ or } \Lambda_2' = 0. \tag{I.4}$$

Now (I.4) has the convenient boundary conditions $\Lambda_1(x) = x$, $\Lambda_2(y) = y$ so that at values of $(t_1, t_2)$ where both $\Lambda_1(t_1) \neq 0$ and $\Lambda_2(t_2) \neq 0$, the unique solution of (2.2) can be obtained from the unique solution of (I.4) by a change of variable. Since $H(t_1, t_2)$ is continuous, the solution can be extended to the closure of the set on which $\Lambda_1'(t_1) > 0$ and $\Lambda_2'(t_2) > 0$. The remaining difficulty occurs when $\Lambda_1'$ or $\Lambda_2'$ equals zero on an interval. Assume $\Lambda_1' = 0$ on $[a, b]$. Then it can easily be shown that $H_{12} = 0$ on $[a, b] \times [0, \infty]$ so that $H(t_1, t_2) = F(t_1) + G(t_2)$ for some $F$, $G$ on this rectangle. A continuity argument then establishes that $H(t_1, t_2) = H(a, t_2)$ on $[a, b] \times [0, \infty]$, and the general solution can be patched together accordingly.

## APPENDIX II

*Some important functions*

For the bivariate exponential form of the model, the contribution to the likelihood of an observation $(x_1, x_2)$ with corresponding censoring indicators $(d_1, d_2)$ is

$$\phi(x_1, x_2, d_1, d_2; \gamma) = (1+\gamma)^{d_1 d_2} \frac{e^{\gamma(x_1 d_1 + x_2 d_2)}}{\{e^{\gamma x_1} + e^{\gamma x_2} - 1\}^{(d_1 + d_2 + 1/\gamma)}}$$

so that the function $\psi$ of Section 4.5 is

$$\psi(x_1, x_2, d_1, d_2; \gamma) = d_1 d_2 \log(1+\gamma) + (1 + \gamma d_1) x_1 + (1 + \gamma d_2) x_2$$
$$- (d_1 + d_2 + 1/\gamma) \log\{e^{\gamma x_1} + e^{\gamma x_2} - 1\}.$$

The first derivatives of $\psi$ with respect to $x_k$, $k = 1, 2$ are required in the calculation of bar-bar scores by methods 1 and 2 and are given by

$$\psi_k = (1 + \gamma d_k) - (1 + \gamma d_1 + \gamma d_2) \frac{e^{\gamma x_k}}{e^{\gamma x_1} + e^{\gamma x_2} - 1}.$$

For the bivariate Pareto form of the model introduced in Section 4.5 and used in Sections 5 and 6, these functions become, for an observation $(y_1, y_2)$ with associated censoring indicators $(d_1, d_2)$,

$$\phi(y_1, y_2, d_1, d_2; \gamma) = (1+\gamma)^{d_1 d_2} \{1 + \gamma y_1 + \gamma y_2\}^{-(d_1 + d_2 + 1/\gamma)},$$

$$\psi(y_1, y_2, d_1, d_2; \gamma) = d_1 d_2 \log(1+\gamma) - (d_1 + d_2 + 1/\gamma) \log(1 + \gamma y_1 + \gamma y_2) + (y_1 + y_2)$$

and,

$$\psi_k = 1 - \frac{1 + \gamma d_1 + \gamma d_2}{1 + \gamma y_1 + \gamma y_2}.$$

## APPENDIX III

*Covariance matrix of scores*

Both approaches to the calculation of approximations to the conditional expectations $\bar{\bar{x}}$ may also be used to approximate the conditional covariance matrix $\Sigma$ used in the variance estimator of 4.2. For our first approach

$$\Sigma_{kl}^{ij} \equiv \operatorname{cov}(x_k^i, x_l^j \mid R_1, R_2) = \sum_{X_k^u \leqslant X_k^i} \; \sum_{X_l^v \leqslant X_l^j} \frac{D_k^u D_l^v}{N_k^u N_l^v} M_{kl}^{uv}$$

where the $2N \times 2N$ matrix $M$ is given by

$$M = W^{-1},$$

$$W_{kl}^{ij} = \delta_{kl}^{ij} - \frac{D_k^i D_l^j}{N_k^i N_l^j} \sum_{\substack{X_k^u \geqslant X_k^i \\ X_l^u \geqslant X_l^j}} \bar{\psi}_{kl}^u$$

where

$$\delta_{kl}^{ij} = 1, \; i = j \text{ and } k = l,$$
$$= 0, \; \text{otherwise},$$

and $\bar{\psi}_{kl}^i$ are as in (4.15), i.e. the second derivatives of $\psi(x_1, x_2, d_1, d_2; \gamma)$ with respect to $x_1, x_2$ evaluated at $(\bar{x}_1^i, \bar{x}_2^i, D_1^i, D_2^i)$.

Our second approach leads to a computationally simpler form which may be shown to be equivalent to an approximation of the above expression using the first two terms of a Neumann expansion for the inverse of $W$.

DISCUSSION OF PAPER BY MR CLAYTON AND DR CUZICK

**Dr Richard D. Gill** (Centre for Mathematics and Computer Science, Amsterdam): This paper is a first really successful attempt at statistically modelling dependent survival times. The authors present a model and accompanying statistical analysis which can be applied to censored matched pairs, to the study of association, and to problems of unobservable covariates. Though the model itself has been around for some time, and despite its close relationship to Cox's proportional hazards model, it has so far resisted satisfactory treatment. The present authors in fact need unfamiliar statistical principles: their methods are based on approximations of marginal likelihoods. A second approximation involved is that the marginal likelihood being used is that appropriate to Type II censoring, even if the actual censoring mechanism is another one.

A lot remains to be done. The methods seem powerful and easy enough for wide application. However, they are only supported by heuristic mathematics, by simulation results and by practical experience.

I would like to suggest an alternative scheme for estimation and testing in these models, based on completely opposed principles: (non-parametric) maximum likelihood using the full likelihood of the observed data. However, there are many striking similarities. To illustrate this approach consider one of the simplest applications: the bivariate symmetric case.

Suppose we observe $(X_k^i, D_k^i)$; $k = 1, 2, \quad i = 1, \ldots, n$ where

$$X_k^i = \min(T_k^i, C_k^i), \qquad D_k^i = I\{T_k^i \leqslant C_k^i\},$$

and $(T_1^i, T_2^i, Z^i)$, $i = 1, \ldots, n$, are i.i.d. triples. $Z^i$ is gamma $(\gamma^{-1}, \gamma^{-1})$ distributed and conditional on $Z^i = z^i$, $T_1^i$ and $T_2^i$ are independent, each with hazard rate $z^i \lambda^0(t)$, $t \geqslant 0$. The $C_k^i$'s are censoring variables. Let $\Lambda^0(t) = \int_0^t \lambda_0(s)\, ds$ be the underlying cumulative hazard rate.

Under "noninformative censoring" (see Arjas and Haara, 1984) the likelihood for the observed data is

$$\prod_{i=1}^{n} \int_{z^i=0}^{\infty} f(z^i;\gamma^{-1},\gamma^{-1}) \prod_{k=1}^{2} z^i \left(\frac{d\Lambda^0}{d\mu}(X_k^i)\right)^{D_k^i} \exp(-z^i\Lambda^0(X_k^i))\, dz^i. \qquad (1)$$

Here $f(z;k,\eta)$ is the gamma density with shape and scale parameters $k$ and $\eta$ respectively and (for the time being) $\mu$ is Lebesgue measure so that $(d\Lambda^0/d\mu)(t) = \lambda^0(t)$.

It is natural to estimate the parameters $\gamma$ and $\Lambda^0(\cdot)$ by maximum likelihood in the spirit of Johansen's (1983) treatment of the proportional hazards model. We propose taking (1) as likelihood for $(\gamma, \Lambda^0)$ even when $\Lambda^0$ is not absolutely continuous, letting $\mu \gg \Lambda^0$ now be arbitrary. The MLE $(\gamma, \hat{\Lambda}^0)$ is the parameter value which maximizes (1) in any pairwise comparison, choosing $\mu$ anew in each comparison. $\hat{\Lambda}^0$ turns out to be a step-function with jumps at the observed failure times only. So we maximize (1) over such $\Lambda^0$'s only, which is the same as maximizing (1) with $(d\Lambda^0/d\mu)(t)$ replaced by the jump at $t$ $\Delta\Lambda^0(t) = \Lambda^0(t) - \Lambda^0(t-)$.

To avoid integrating out the $z^i$'s we may use the *EM* algorithm. Suppose for simplicity $\gamma$ is known. Since $z_i$ enters in (1) as the gamma $(\gamma^{-1} + \Sigma_k D_k^i, \gamma^{-1} + \Sigma_k \Lambda^0(X_k^i))$ density, the algorithm has as *E*-step:

$$\hat{Z}^i = \frac{\gamma^{-1} + \Sigma_k D_k^i}{\gamma^{-1} + \Sigma_k \hat{\Lambda}^0(X_k^i)}$$

and *M*-step:

$$\Delta\hat{\Lambda}^0(t) = \frac{\Sigma_{i,k} D_k^i I\{X_k^i = t\}}{\Sigma_{i,k} \hat{Z}^i I\{X_k^i \geq t\}}.$$

These are simply the bar-bar iterations (4.17).

Since the integral in (1) has a simple closed form expression, estimation of $\gamma$ is most easily done by maximizing (1) over $\Lambda^0$ for each of a range of values of $\gamma$.

What about justification of this method? Again, only heuristics are available. If these estimates are a smooth enough function of the data, very nice large sample properties follow immediately: asymptotic normality and even efficiency in the sense of Begun *et al.* (1983); Gill, (1985); and Bickel (personal communication). Note that we need to extend (1) as smoothly as possible to discrete $\Lambda^0$'s, not as realistically as possible. However the "if" here is a very hard problem.

So both approaches are very close and proving their large sample properties just as difficult. We hope others will take up the challenge too.

**Professor P. Armitage** (University of Oxford): The authors are to be congratulated on laying the foundation of a general theory of association for right-censored variables, and incidentally for managing to reconcile the different approaches they had previously taken. I am glad that they refer to applications in the econometric literature. The analysis of survival data has mainly been associated with medical applications, but right-censored time variables occur in the social sciences, in biology and industrial technology, and the topic should be regarded as part of the general methodology of statistics.

The authors describe a measure of association playing a role analogous to that of the correlation coefficient in bivariate normal data. I was reminded of the remarks made by F. J. Anscombe in the discussion on Hotelling (1953). "It seems to me", said Anscombe, "that in most multivariate problems when correlation coefficients are calculated, they are of no interest in themselves but only as a step in calculating something else. . . . Am I right in suggesting that the cases where a correlation coefficient is itself of direct interest are rather rare?" A similar point could be made about the topic under discussion today. The question at issue with bivariate survival times is often one of regression. If a cancer patient has a given remission time, what can be said about the distribution of subsequent survival? The authors allude to these questions, but it would be useful if they could explain more clearly the connection between the correlation and regression aspects of their theory.

The "frailty" concept inevitably brings to mind the concept of "accident proneness", and the two approaches to the basic model, outlined in Section 2, are reminiscent of the way in which bivariate accident distributions can arise by mechanisms other than proneness. One of the motivations of proneness theory was the hope that the removal of people with high initial accident experience would substantially lower the subsequent incidence. On the whole this hope was illusory. Similar questions might arise with bivariate survival data. If individuals with low values of $T_1$ were removed, how would that affect the marginal distribution of $T_2$? Has this been considered by the authors?

Finally, nomenclature. The authors have used a new term, "bar bar scores", conjuring up visions of forty thieves, black sheep and the king of the elephants, presumably in a bid to capture the under-five market. In my view the authors will need to simplify their mathematics if they hope to do this, and I look forward to their next move in this direction.

It gives me pleasure to second the vote of thanks.

The vote of thanks was carried by acclamation.

**Dr J. R. Whitehead** (University of Reading): This ingenious paper has introduced an analogue of the correlation coefficient into the study of bivariate censored data. However, in the analysis of ordinary data, I tend to be a regressor rather than a correlator. Therefore, I am interested in how the approach of tonight's authors relates to a proportional hazards regression approach.

Some of the examples mentioned in the introduction seem to be well suited to regression modelling. When considering the effect of time from remission to relapse on time from relapse to death, the former time will be uncensored and usually a prediction of the latter will be required. Again, if the heart-attack-free lifetime of a father is to be used as a predictor of the same variable for his son, then a regression model, with a conclusion in terms of relative hazard would seem appropriate. However, now there is a problem: the father's heart-attack-free lifetime is a censored observation.

One way of modelling such a covariate is via the linear term

$$\beta_1 a + \beta_2 h,$$

which would appear in the exponent of the relative risk. Here $a$ is the number of years survived by the father without a heart attack, and $h$ equals 1 if father had a heart attack and 0 otherwise.

Now consider two men; the father of Case 1 had a first heart attack at age $a$, the father of Case 2 has had no heart attacks and is age $a$. Thus, relative to some baseline hazard function $\lambda_0(t)$ the men have the hazard functions

$$\lambda_1(t) = \exp(\beta_1 a + \beta_2)\,\lambda_0(t)$$

$$\lambda_2(t) = \exp(\beta_1 a)\,\lambda_0(t)$$

respectively. Hence

$$\lambda_1(t) = \exp(\beta_2)\,\lambda_2(t).$$

From Section 2 of tonight's paper we see that the authors' $\theta$ is equal to $\exp \beta_2$. Of course $\exp \beta_2$ can be estimated from the sons' survival times by ordinary proportional hazards regression.

Suppose, next, that the dependence of the fathers' hazards on their sons' heart-attack-free lifetime can be modelled in the same way, using the linear term $\gamma_1 a + \gamma_2 h$. Now $a$ and $h$ refer to the son. Section 2 also shows that $\theta = \exp \gamma_2$. The paper shows how the fathers' survival times can be combined with those of the sons to provide a more accurate means of estimating $\theta$. This greater accuracy is only present if both dependencies can be modelled as above, and this would require careful checking. Presumably, the linear model methods of Section 5 could be used to answer questions about $\beta_1$ and $\gamma_1$, as well as incorporating other covariates into the analysis.

**Professor Murray Aitkin** (University of Lancaster): The model (5.1) with a shared random effect in the log hazard is a generalization of an exponential family model with a random effect in the linear predictor. These models are widely useful because they can represent a wide range of "extra variation": heterogeneity due to omitted explanatory variables, variance components for several levels of nesting, measurement error in the explanatory variables, or outlying observations. Finding parametric models for the random effects which integrate to a tractable form for the marginal distribution of the observed data is both difficult and restrictive.

If we *estimate* the distribution of the "mixing variable" by nonparametric maximum likelihood, we avoid completely the specification of this distribution, and gain a very powerful and general procedure based on fitting finite mixture distributions, which can be achieved using *GLIM* macros.

The approximate marginal likelihood approach forced on the authors by using rank methods is quite difficult to handle, though the authors are ingenious in their development of *EM* algorithmic approaches recognizing the order statistics as missing data and the ranks as observed data.

A possible alternative semi-parametric approach to the model (5.1) is to reverse the authors' approach: to specify a full parametric kernel distribution for the $\epsilon_i$ *without* the monotone transformation, but to estimate nonparametrically the distribution of $w$ in fitting the model. This is quite straightforward computationally.

**Dr A. N. Pettitt** (University of Loughborough): I congratulate the authors on a comprehensive and stimulating paper. I wish to make these points.

(i) The authors refer to Pettitt (1982) where approximate estimates for the model of Section 5.1 are given. This approximate estimate of $\beta$ is a 1-step Newton–Raphson estimate starting from $\beta = 0$ using the rank likelihood. Standardizing the explanatory variable $z$ to lie in $[-1, 1]$, it is straightforward to show that this approximate estimate is asymptotically efficient, in terms of the rank likelihood, if $\beta = o(1)$; see Cox and Hinkley (1974, p.308). This allows for $\hat{\beta}/(\text{var}\,\hat{\beta})^{1/2} \to \infty$, that is the $P$-value for $H$: $\beta = 0$ can go to zero.

(ii) A fully parametric model might be preferred if predictions about individual observations are required or the transformation, $g$, say, to the canonical model can be identified with high precision. For original observations on $Y$ giving the canonical model

$$g(Y) = \beta^{\mathrm{T}} z + \text{error}$$

the authors' method of estimation provides approximations to $E(g(Y) \mid \text{ranks}, \beta)$, that is the $\bar{y}$ or $\bar{\bar{y}}$ scores. If the original observations are plotted against their final $\bar{\bar{y}}$ scores then this gives a curve close to $g^{-1}(\cdot)$, so trial, transformations of the $Y$'s can be plotted against these scores, seeking a straight line relationship. Note, of course, that the $g(\ )$ function can change as explanatory variables are removed or added to the model.

(iii) The expectations $E(g(Y) \mid \text{ranks}\ \beta)$ at the final estimate $\hat{\beta}$ are also useful in the assessment of influential cases, in the spirit of Cook and Weisberg (1982).

(iv) Without specification of the transformation, how can these models be used to relate to the original observations or future observations? Pettitt (1982, Section 5) and Pettitt (1983, Section 7.1) made some suggestions. For example, for the sibling model of Section 6, within sibship comparisons can be made by considering Pr $(S_i > S_{i'} \mid \beta)$. Now

$$S_i - S_{i'} = \beta^{\mathrm{T}}(z_i - z_{i'}) + \epsilon_i - \epsilon_{i'}$$

and the event $\{S_i > S_{i'}\}$ is invariant to monotone increasing transformation. Pettitt (1983) also considered the maximum median response for a quadratic response model. Another approach, Acar and Pettitt (1984), considers the probability that a future observation has a value between the observations having ranks $j$ and $j + 1$.

The following contributions were received in writing, after the meeting

**Dr S. M. Gore** (MRC Biostatistics Unit, Cambridge): Mr President, ladies and gentlemen, I congratulate Mr Clayton and Dr Cuzick on a very fine collaboration, which was initiated three years ago on the snow-covered hills of the Black Forest. This evening's paper is a fitting tribute to the Oberwolfach ideal of intellectual and social exchange in an exhilarating, isolated and beautiful location.

Mr Clayton and Dr Cuzick generalize the proportional hazards model to the problem of two time variables with unspecified marginal distributions which are related by a single association parameter. The eccentricity of the proportional hazards model within this generalized structure is noteworthy; accelerated failure time models, such as the log-logistic, are accommodated neatly within this family of models characterized by proportionality of conditional hazard functions, conditional, that is, on unmeasured frailty $\omega$, for which gamma or inverse Gaussian distribution is a usual assumption.

What further generalization of this structure is practical? Negative correlation, as might obtain in respect of alcoholism in families, needs investigation; the link between a generalized accelerated failure time model and time variables whose unspecified marginal distributions are related proportionately to *two* association parameters might be explored. A different tack is to consider time-dependence of the association parameter, $a_k$. A step function is the simplest example and could be appropriate when there is some distinction between early and late onset of disease, as in diabetes. In litter-matched studies, the strength of association with the unobserved covariate $\omega$ might depend on birth weight or birth order, so that the exchangeability implied by setting $a_k = 1$ for all $k$ is thwarted.

Mr Clayton's and Dr Cuzick's paper repays repeated reading – the last paragraph in Section 5.2, for example, contains three tests of association between two dependent variables. These are for consideration when intra-class correlation is confounded with heterogeneity of frailty, as in genetic/environmental controversies.

The numerical results at the end of the paper do not deal with misspecification of the integrated conditional hazard. How robust are the authors' proposed tests? Am I correct in surmising that all methods, parametric as well as bar-scores, derived from rank order information overestimate gamma when sample size is small (less than 100) or censoring heavy (10 to 50 per cent)? More detailed Monte Carlo studies would be instructive. But when the authors have given so much, it is ungrateful to ask for more!

**Dr R. Kay** (University of Sheffield): It is rare nowadays to find a paper in survival analysis that is innovative. This however, I'm pleased to say, is an exception and Mr Clayton and Dr Cuzick are to be congratulated on producing such a novel approach to the analysis of bivariate survival data. Some questions, however, remain to be answered particularly in relation to the interpretation of analyses based on these techniques. The authors mention in the introduction several examples of bivariate data which may lend themselves to this kind of correlation analysis. One such example (". . . in cancer trials it is of interest to know if the interval from remission to relapse influences the subsequent interval from relapse to death.") concerns the dependence of an inter-event time on another inter-event time in a stochastic process. Problems of this type have already been tackled using proportional hazard models with inter-event times for preceding events included in the covariate vector of the model.

References are Aalen *et al.* (1980), Kalbfleisch and Prentice (1980) Section 7.3 and Kay (1982), (1984). For example suppose $z_1$ is the time for remission to relapse and $z_2$ is the time from relapse to death. The proportional hazards model assumes that the hazard function for $z_2$ for a patient with covariate vector $\mathbf{x}$ is

$$\lambda(z_2 \mid \mathbf{x}, z_1) = \lambda_0(z_2)e^{\boldsymbol{\beta}'\mathbf{x}+\alpha z_1}.$$

Alternative definitions of the time origin are possible.

Partial likelihood (Cox, 1975) methods produce estimates of $\alpha$ and $\boldsymbol{\beta}$ and a large sample test of $H_0: \alpha = 0$ (no association) is easily produced. Such models are straightforward to interpret and can be fitted very easily using standard survival data programs such as *BMDP*2. What is the connection between this approach and that proposed by Cuzick and Clayton? Similar methods, which involve modelling in terms of one time scale while including alternative time scales in the proportional hazards covariate vector, have been used in a wider context by Farewell and Cox (1979).

**Professors Ross L. Prentice and Steven G. Self** (Fred Hutchinson Cancer Research Center, Seattle, Washington, USA): We are pleased to comment on topics related to this interesting paper. First we would like to note that a score test for independence of the bivariate failure times can be derived from the authors' expressions $\lambda_k(t \mid w)$, $k = 1, 2$ at the beginning of Section 2 without introducing the additional assumptions of their bivariate proportional hazards model (i.e. without assuming $w$ to have a log gamma distribution). Specifically, without loss of generality we may assume the frailty $w$ to have mean zero and variance $\sigma^2$. Denote $\epsilon = w\sigma^{-1}$ and suppose that $w$ has density $f$. A score test for independence can now be written

$$\lim_{\sigma \downarrow 0} \partial \log L/\partial\sigma^2 = \lim_{\sigma \downarrow 0} \tfrac{1}{2} \frac{\partial^2 \log L}{\partial\sigma^2},$$

where $L$ is the marginal likelihood function

$$L = \int \prod_{i=1}^{n} \prod_{k=1}^{2} \left\{\exp(a_k \sigma \epsilon_i) \Big/ \sum_{l \in R_k(X_k^i)} \exp(a_k \sigma \epsilon_l)\right\}^{D_k^i} \prod_{j=1}^{n} f(\epsilon_j)\, d\epsilon_j,$$

and where $R_k(t)$ denotes the sample $k$ risk set at time $t$, $k = 1, 2$. Straightforward calculations show this score statistic to be proportional to

$$T = \sum_{i=1}^{n} \{\hat{\Lambda}_1(X_1^i) - D_1^i\}\{\hat{\Lambda}_2(X_2^i) - D_2^i\},$$

where

$$\hat{\Lambda}_k(t) = \sum_{\{j \mid X_k^j \leqslant t\}} D_k^j n_{kj}^{-1},$$

and $n_{kj}$ is the size of the sample $k$ risk set at $X_k^j$. This is the same score statistic alluded to in Section 3. Simple counting process arguments yield

$$\sum_{i=1}^{n} \prod_{k=1}^{2} \left[ \sum_{\{j \mid X_k^j \leqslant X_k^i\}} D_k^j n_{kj}^{-1} (1 - D_k^j n_{kj}^{-1}) \right]$$

as a variance estimator for $T$ under quite general censorship.

The same approach can be used to generate a score test of independence in the presence of additional covariates, and an appropriate variance estimator can be developed. In general the additional assumption of equality of frailty coefficients ($a_1 = a_2$) is required to eliminate a dependence of the standardized score test on $\{a_1, a_2\}$.

In the presence of dependence between $T_1$ and $T_2$ the authors propose simultaneous estimation of their association parameter $\gamma$ (Section 4) or simultaneous estimation of $\gamma$ and regression parameters (Section 5) within the context of their rather specialized bivariate model. A rather more empirical and more flexible approach would directly model the dependence of the failure rate for subject $(k, i)$ at time $t$ on the preceding failure time history for both members of the $i$th pair. Assuming independent censorship and independent pairs of failure times the intensity process for subject $(k, i)$ can be written

$$\lambda_{ki}(t) = Y_{ki}(t)\, \lambda_k\{t \mid z_{ji}, N_{ji}(u^-), Y_{ji}(u), u \leqslant t, j = 1, 2\}$$

where $Y_{ki}(t) = 1$ if subject $(k, i)$ is at risk at time $t$ and equals zero otherwise, while $N_{ki}$ is the failure time counting process for subject $(k, i)$. As a simple example, one may specify a relative risk type model

$$\lambda_k\{t \mid z_{ji}, N_{ji}(u^-), Y_{ji}(u), u \leqslant t, j = 1, 2\} = \lambda_{0k}(t) \exp\{z_{ki}\beta_k + N_{li}(t^-)\gamma_k\}, k = 1, 2; l \neq k.$$

This model specifies a hazard rate $\lambda_{0k}(t) \exp\{z_{ki}\beta_i\}$ for subject $(k, i)$ up to the time of failure of the other pair member, at which time the hazard rate becomes $\lambda_0(t) \exp\{z_{ki}\beta_k\}\, e^{\gamma_k}$. Standard partial likelihood methods can be used both for testing independence of failure time pairs ($\gamma_1 = \gamma_2 = 0$) and for regression parameter estimation in the presence of association. This method would readily extend more complex relative risk models with time-dependent covariates and non-proportional dependencies on $N_{li}(t^-)$. Note, however, that some modification of the $(k, i)$ intensity model would likely be appropriate at times after the corresponding $T_l^i (l \neq k)$ has been censored.

We would be interested in the authors' views of these alternatives to their proposals.

**Dr Philip Hougaard** (Novo Research Institute, Bagsvaerd, Denmark): First I would like to thank the authors for this stimulating paper containing many ideas which hopefully will be followed up in the future. However, I don't think that the assumption of gamma distributed frailties is convenient in the multivariate situation with covariates. As the authors describe in Section 5.2 the association is confounded to population heterogeneity. This means that the joint distribution of $(T_1, T_2)$ can be identified from the marginal distributions and thus the association parameter describes more than just association. This problem is present for all frailty distributions with finite mean. Instead one can assume that the frailties follow a (standardized) positive stable distribution with index $\alpha \in (0.1]$ given by the Laplace transform $E \exp(-sX) = \exp(-s^\alpha)$. Some results about these distributions are in Hougaard (1984), which considers frailty distributions from a three-parameter family containing the positive stable distributions, the gamma distributions and the inverse Gaussian distributions. With a positive stable distribution the formula corresponding to that in Theorem 1 in Section 2 becomes

$$H(t_1, t_2) = \{\Lambda_1(t_1)^{1/\alpha} + \Lambda_2(t_2)^{1/\alpha}\}^{\alpha}$$

Expression (6.1) describes a model, where conditionally on the $\omega$'s the hazards are proportional with factors $\exp\{\beta^{\mathrm{T}}(Z_2 - Z_1) + \omega_2 - \omega_1\}$. With gamma distributed frailties the hazard ratios in the marginal distributions converge to 1, but with a positive stable distribution the hazards in the marginal distributions are proportional with factors $\exp\{\alpha\beta^{\mathrm{T}}(Z_2 - Z_1)\}$. A paper on how the positive stable distributions can be applied to multivariate life-tables is in preparation.

The **authors** replied later, in writing, as follows.

We would like to thank the discussants for their contributions which raise several interesting and important points. Several contributions deal with related issues and we shall first deal with these more general ideas before answering more specific points.

Professor Armitage draws an analogy with correlation and regression in the bivariate normal case, and similar ideas are implicit in the contributions of Drs Whitehead and Kay and that of Professors Prentice and Self. Dr Kay refers to the case where we are primarily interested in the regression of $t_2$ on $t_1$, but, importantly, $t_1$ is uncensored. In such a case we have no need for a bivariate model and, as Dr Kay notes, we may use a variant of the proportional hazards model to describe the distribution of $t_2$ conditional upon $t_1$. However the problems with which we are mainly concerned involve right censored observations of $t_1$ and, as in any problem in which covariates are imperfectly measured, it becomes necessary to adopt a bivariate model. Dr Whitehead makes an interesting attempt to avoid this necessity by introducing the censored covariate, $t_1$, as two covariates—the last *observed* age and the censoring indicator. However, this introduces two parameters to describe the relationship and there is some difficulty in their interpretation. When $t_1$ is heavily censored the censoring indicator contains most of the information relevant to the relationship between $t_1$ and $t_2$, but when $t_1$ is uncensored we have the situation described by Dr Kay where the observed values of $t_1$ contain the information. In intermediate cases, both $\beta_1$ and $\beta_2$ measure the association and it is not clear how one might interpret them. It must be conceded, though, that if we are not primarily interested in the effect of $t_1$ but wish only to account for its effect when investigating other factors in a regression analysis for $t_2$ then Dr Whitehead's suggestion is attractive and simple.

If we must consider a bivariate model, then it is clearly desirable that the model we choose has a regression interpretation. Indeed, this was one of the motivating considerations set out by Clayton (1978) when first proposing the model discussed in our paper. Similar considerations have been discussed by Cox and Oakes (1984, Chapter 10). The regression interpretation of the model is essentially contained in the relationship

$$\lambda_2(t_2 \mid T_1 = t_1) = (1 + y)\, \lambda_2(t_2 \mid T_1 > t_1).$$

From our Theorem 1, the hazard function for $t_2$ conditional upon $(t_1, d_1)$ may be represented in terms of the marginal hazard and integrated hazard functions by

$$\lambda_2(t_2 \mid t_1, d_1) = (1 + \gamma)^{d_1} \frac{\exp \gamma\Lambda_2(t_2)}{(\exp \gamma\Lambda_1(t_1) + \exp \gamma\Lambda_2(t_2) - 1)} \lambda_2(t_2). \qquad (1)$$

The middle term depends upon $t_2$ so that this is not, in general, a proportional hazards model. Incidentally, this term represents the factor by which the marginal hazard, $\lambda_2(t_2)$, is reduced by removal of frail individuals (by deleting those with small $T_1$ values). This is the factor about which Professor Armitage enquires; for constant $t_1$ it attains its minimum value at $t_2 = 0$ and increases monotonically to unity as $t_2$ increases to infinity. Thus, equation (1) is a convergent hazards model. However, in the epidemiological situations which motivated the model, $\Lambda_1(t_1)$ and $\Lambda_2(t_2)$ are both small and this term may be closely approximated by $\exp - \gamma\Lambda_1(t_1)$. In these circumstances, (1) may be approximated by the proportional hazards model

$$\lambda_2(t_2 \mid t_1, d_1) \cong \lambda_2(t_2) \exp\{-\gamma\Lambda_1(t_1) + d_1 \log(1 + \gamma)\}.$$

Note that, if $\Lambda_1(t_1)$ is linear in $t_1$, we obtain Dr Whitehead's model.

Other approaches involve the assumption of different bivariate models. There are, of course, several possible models with attractive features. It is not clear to us why the model discussed by Professors Prentice and Self is "more empirical and more flexible" than ours or, indeed, any less specialized. The model they advocate is essentially a generalization of the bivariate exponential distribution of Freund (Freund, 1961; Johnson and Kotz, 1972) and does seem more appropriate in situations in which *the occurrence* of one event directly influences a related failure process. However, in other situations, this model is distinctly unattractive. For example, in the context of familial association of disease incidence, the Freund model would predict that a son's hazard function would follow one curve up to the age at which his father contracted the disease and then jump to another curve at a higher level of risk. This hardly seems biologically plausible. Also, the problem of right censoring of the covariate (father's age at failure) is unresolved since after the age of loss to follow-up of the father, it is not known which hazard curve the son is following. This raises a problem akin to the assumption of a two-point frailty distribution and its solution would probably require an *EM* approach along similar lines to that which we have proposed.

We were most interested in the alternative bivariate model proposed by Dr Hougaard. This distribution has been described before (Johnson and Kotz, 1972) and has been termed a "bivariate extreme value distribution", but its interpretation as a heterogeneous frailty model in which the proportional hazards model for covariates is not destroyed in the margins is both original and important. We do not have any intuition as to the plausibility of the positive stable distributions for frailty and some sketches for realistic values of $\alpha$ would be of interest. Against this model it can be said that no regression interpretation similar to that discussed above seems to be available. Also, as Dr Gore has pointed out, the destruction of the proportional hazards model by the gamma frailty assumption is interesting and potentially useful in its own right. The semi-parametric Pareto regression model described in Section 5.2 may be regarded as a super-model which includes both proportional hazards and proportional odds models as special cases.

Dr Gore also raises some questions concerning the possible extension of our model. Certainly its inability to model negative association is a limitation. Such association might arise as a result of competition and might be approached by a model in which the logarithm of frailities (or some other transform of frailties) sum to a constant value within a family. The possibility of time-dependent frailty effects is interesting as are further "interaction" generalizations such as random susceptibilities to fixed covariates. Much remains to be done before we obtain a fully generalized proportional hazards model with a variety of fixed and random effects. The search for such a model is difficult owing to the difficulty, to which Professor Aitkin draws attention, in finding more general frailty distributions with sufficiently convenient analytical properties. We have carried out further work on the multivariate gamma frailty model to which we refer in our paper and would hope to present this work soon. It does not however offer solutions to all the problems described above, and one is hopeful that the work referenced by Dr Hougaard will lead to further generalizations.

We now move on to discuss the points concerning our method of estimation. Professor Gill is quite right to draw attention to the need for further more rigorous work. Although we have more recently developed a new approach to the calculation of scores which more formally justifies method 2 of this paper, the total problem remains difficult. Dr Pettitt and Professor Gill both note that our computation of bar-bar scores is akin to estimation of the baseline hazard functions $\Lambda_1^0$ and $\Lambda_2^0$. We would agree with Dr Pettitt that these functions are of considerable interest in making predictions from the model; indeed our discussion of the regression interpretation of the model clearly demonstrates the need for knowledge of the form of these functions. We made reference in the paper to the fact that estimation of $\Lambda_k^0$ by step functions with discontinuities at the observed failure times (as suggested by Breslow for the proportional hazards model) leads to our method with scores calculated by method 2. Indeed, between ourselves we have been referring to these scores as "Breslow scores". We are grateful to Professor Gill for a more formal description of our model within the counting process framework. However, as he states, the justification of asymptotic properties using this approach also remains, for the present, heuristic. We would join him in commending the problem to other workers.

Professor Gill's remarks also suggest a line of argument which clarifies our approximate likelihood. To ease the notation we shall consider the univariate regression case of Section 5.2. Conditional upon frailities, $\xi$, the parametric kernel $L_k(y, \beta, \xi)$ is a product of exponential densities and survivor functions. The marginal likelihood is

$$L_M = \int_{y \in R} \int_{\xi} L_k(y, \beta, \xi)\, dG(\xi; \gamma)\, dy$$
$$= \int_{\xi} \int_{y \in R} L_k(y, \beta, \xi)\, dy \quad dG(\xi; \gamma)$$

where

$$dG(\xi; \gamma) = \prod_i g(\xi_i; \gamma)\; d\xi_i.$$

The inner term is a partial likelihood and, using the approach of Johansen (1983), it may be shown to be identical to the maximum with respect to $y$, $y \in R$ of a total likelihood of the form $L_k(y, \beta, \xi)\, J(y)$ where $J(y)$ is the Jacobian-like term given by the product of the spacings, $(y_i - y_{i-1})$, of the transformed uncensored times. Thus

$$L_M = \int_{\xi} \max_{y \in R} \{L_k(y, \beta, \xi)\, J(y)\} \quad dG(\xi; \gamma)$$

Professor Gill's remarks, together with our note of the equivalence between our method and that of Bennett for the special case $\gamma = 1$, show that our approximation, with method 2 scores, is equivalent to the use of the approximate likelihood

$$L_M^* = \max_{y \in R} \left[ \int_{\xi} L_k(y, \beta, \xi)\, dG(\xi; \gamma)\, J(y) \right]$$

This is not identical to $L_M$, but our work presented here and later work which more formally justifies the method 2 scores (Clayton and Cuzick, 1985a) suggests that it has the usual asymptotic properties of a likelihood.

Professors Prentice and Self repeat our derivation of score tests for association using the marginal likelihood argument, but give a different variance estimator from that of Cuzick (1982), whose results are also valid for a wide class of error distributions. This latter estimator has also been derived by Dabrowska (unpublished) assuming only that the censoring times $(C_1, C_2)$ are independent of the failure times $(T_1, T_2)$ and of all other sets of $(T_1, T_2, C_1, C_2)$. The conditions under which the variance estimator proposed by Prentice and Self is more appropriate are not clear to us. We note in passing that the simple counting process arguments they invoke only seem useful under the hypothesis of no association. In other cases, it is no longer sufficient to condition simply upon the past and this fact seriously hampers rigorous justification of the asymptotic properties of estimators.

Professor Aitkin suggests an alternative semi-parametric approach to such problems without our arbitrary monotone transformation but with an arbitrary frailty distribution. Although an interesting possibility, it is unlikely that this model is sufficiently flexible for the applications we envisage. For example, all mixtures of exponential distributions have decreasing hazard. Thus, a free choice of frailty distribution leads to a very restricted set of marginal failure time distributions. An interesting possibility is to allow both arbitrary monotone transformation of times and arbitrary frailty distribution. Then our method 2 would become a quasi-*EM* algorithm in which the unknown frailties are replaced by *non-parametric* empirical Bayes estimates at each $E$-step. The theoretical complications of this generalization would appear formidable, however, even in the simple regression case of Section 5.2.

Finally we must apologise if the above remarks do not answer all the points made by the discussants, for which we are most grateful. In particular we must thank Dr Gore for so eloquently reminding us of our debt to the organisers of that Oberwolfach meeting.

## REFERENCES IN THE DISCUSSION

Acar, B. S. and Pettitt, A. N. (1984) Prediction using non-parametric techniques: some comparisons. *J. Statist. Comput. Simul.*, in the press.

Arjas, E. and Haara, P. (1984) A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.*, **11**, 193–209.

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983) Information and asymptotic efficiency in parametric–non-parametric models. *Ann. Statist.*, **11**, 432–452.

Clayton, D. G. and Cuzick, J. (1985c) The semi-parametric Pareto model for regression analysis of survival times. *ISI Bull.*, in the press.

Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. London: Chapman and Hall.

Cox, D. R. (1975) Partial likelihood. *Biometrika*, **62**, 269–276.

Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Farewell, V. T. and Cox, D. R. (1979) A note on multiple time scales in life testing. *Appl. Statist.*, **28**, 73–75.

Freund, R. J. (1961) A bivariate extension of the exponential distribution. *J. Amer. Statist. Ass.*, **56**, 971–977.

Gill, R. D. (1985) Non- and semi-parametric maximum likelihood estimators and the von Mises method. Preprint MS-R8509, Centre for Mathematics and Computer Science, Amsterdam, Netherlands.

Hougaard, P. (1984) Frailty models derived from the stable distributions. Preprint 84/7, Institute of Mathematical Statistics, University of Copenhagen, Denmark.

Johansen, S. (1983) An extension of Cox's regression model. *Int. Statist. Rev.*, **51**, 165–174.

Johnson, N. L. and Kotz, S. (1972) *Continuous Multivariate Distributions*. New York: Wiley.

Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kay, R. (1982) The analysis of transition times in multistate stochastic processes using proportional hazard regression models. *Commun. in Statist., Theory & Methods*, **11**, 1743–1756.

—— (1984) Multistate survival analysis: an application in breast cancer. *Meth. Inf. Med.*, **23**, 157–162.

Pettitt, A. N. (1983) Approximate methods using ranks for regression with censored data. *Biometrika*, **70**, 121–132.

As a result of the ballot held during the meeting, the following were elected Fellows of the Society.

Brodie, Judith A.
Collis, Glyn M.
Flowerdew, Gordon
Friday, Dennis S.
Garrett, Andrew D.
Jakeman, Nicola
Katz, Barry P.
Murray, Clive L.
Perry, Joseph N.
Wafula, Charles
Walloe, Lars
Bales, Kevin B. L.
Ganeshanadam, Sivapathasundaram
Gluntz, Carlo J.
Hamilton, Carolyn B. J.
Jones, Anthony R.
Lindsey, Jane C.
Mar-Molinero, Cecilio
Paton, Alexander G.
Shimpi, Prakash A.
Tang, San-Yee
Walters, David E.
Whittington, John R.
Wormser, Wolf E.