Lifetime Data Analysis, 8, 5–19, 2002

Optimal Weight Functions for Marginal Proportional Hazards Analysis of Clustered Failure Time Data

ROBERT J. GRAY gray@jimmy.harvard.edu
Department of Biostatistical Science, Dana-Farber Cancer Institute, Boston, MA 02115, USA

YI LI yili@jimmy.harvard.edu
Department of Biostatistical Science, Dana-Farber Cancer Institute, Boston, MA 02115, USA

Received September 15, 2000; Revised April 3, 2001; Accepted June 27, 2001

Abstract. The choice of weights in estimating equations for multivariate survival data is considered. Specifically, we consider families of weight functions which are constant on fixed time intervals, including the special case of time-constant weights. For a fixed set of time intervals, the optimal weights are identified as the solution to a system of linear equations. The optimal weights are computed for several scenarios. It is found that for the scenarios examined, the gains in efficiency using the optimal weights are quite small relative to simpler approaches except under extreme dependence, and that a simple estimator of an exchangeable approximation to the weights also performs well.

Keywords: Estimating equations, conditional martingale covariance rate, Archimedean copula

1. Introduction

Wei, Lin and Weissfeld (1989) introduced marginal proportional hazards analysis of multivariate failure time data. Their approach was to fit separate marginal proportional hazards models to each failure type and draw inferences using the joint distribution of the separate estimates. In the clustered data setting with common regression effects and baseline hazards across clusters, Lee, Wei and Amato (1992) showed that estimators of the regression parameters obtained from fitting the standard partial likelihood for independent data were consistent. Cai and Prentice (1995) extended this approach to general multivariate failure time data, and proposed modifying the partial likelihood scores by incorporating weight matrices to improve efficiency. Cai and Prentice (1997) considered similar equations for the clustered data setting, and Prentice and Hsu (1997) considered joint weighted estimating equations for marginal regression parameters and association parameters.

Here the clustered data setting of Lee, Wei and Amato (1992) and Cai and Prentice (1997) is considered. Let T_{ij}, C_{ij} and z_{ij} be the failure time, censoring time, and (fixed) covariate vector for subject j in cluster i, $j = 1,\ldots, n_i$, $i = 1,\ldots, N$. We assume the C_{ij} are iid and independent of the T_{ij}, and set $G(t) = P(C_{ij} > t)$. The observed data are right censored with only $X_{ij} = \min\{T_{ij}, C_{ij}\}$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$ observed. The marginal proportional hazards model specifies that the hazard rate for T_{ij}, conditional on covariate information, but not on failure information for other subjects, is $\lambda(t|z_{ij}) = \lambda_0(t)\exp(z_{ij}'\beta)$, where β is a vector of unknown parameters and $\lambda_0(t)$ is an unknown baseline hazard function. Clusters are assumed to be independent, but failure times within a cluster can be correlated. We also assume that the bivariate distribution of any within cluster pair of observations can be written in the form

$$S_{ijk}(s,t) = P(T_{ij} > s, T_{ik} > t) = H\left\{\Lambda_0(s)e^{z_{ij}'\beta}, \Lambda_0(t)e^{z_{ik}'\beta}\right\}, \tag{1}$$

for a symmetric bivariate survival function $H(\cdot,\cdot)$, where $\Lambda_0(t) = \int_0^t \lambda_0(u)\,du$. This model assumes that the dependence on the covariates is only through the marginal proportional hazards model, and that there is no additional structure that would imply different association models for different pairs. Such a model would usually not be appropriate for repeated event data, for example, but might be suitable for modeling center effects in clinical trials. We also note that (1) is equivalent to the standard formulation of joint distributions in terms of copula functions (see Genest and MacKay, 1986, and Marshall and Olkin, 1988), which express the joint distribution in terms of the marginal cumulative distribution functions, since the cumulative hazard Λ and CDF F are related through $\Lambda(x) = -\log\{1-F(x)\}$.

Define the standard counting processes by $N_{ij}(t) = I(X_{ij} \leq t, \delta_{ij} = 1)$ and $Y_{ij}(t) = I(X_{ij} \geq t)$, the marginal martingale process by $M_{ij}(t,\beta) = N_{ij}(t) - \int_0^t Y_{ij}(u)\,\lambda(u|z_{ij})\,du$, and the estimated martingale process by $\hat{M}_{ij}(t,\beta) = N_{ij}(t) - \int_0^t Y_{ij}(u)\exp(z_{ij}'\beta)\,d\hat{\Lambda}_0(u)$, where $\hat{\Lambda}_0(t) = \int_0^t \sum_{i,j} dN_{ij}(u)/\sum_{k,l} Y_{kl}(u)\exp(z_{kl}'\beta)$. Then the family of weighted estimating equations of Cai and Prentice (1997) can be written

$$D(\beta) = \sum_i \sum_j \sum_k \int z_{ik}\omega_{ikj}(t,\beta)\,d\hat{M}_{ij}(t,\beta) = 0, \tag{2}$$

where the $\omega_{ikj}(t,\beta)$ are arbitrary weight functions. The regression parameter estimator $\hat{\beta}$ is defined as the solution to this system of equations.

Ideally, the weight functions would be chosen to minimize the variance of $\hat{\beta}$. However, the form of the optimal weights is not known. Cai and Prentice (1995, 1997) and Prentice and Hsu (1997) restrict attention to constant functions of time in their numerical results, but even within that class the optimal choice of weights was not identified.

In this paper we consider families of weights which are constant on time intervals. We give the explicit form of the large sample variance of $\hat{\beta}$ for such weights, and identify approximately optimal piecewise constant and time-constant weights as solutions to systems of linear equations. Although these weights are not practically useful because of their dependence on the true distribution, it is of interest to compare their performance to simpler approaches to see how much efficiency could be gained. We investigate this question by comparing the performance of estimating equations using the optimal piecewise constant weights, piecewise constant weights that are optimal for ‘exact’ martingale estimating equations, the optimal constant weights, the constant weights optimal for the exact martingale equation, and the independence working model. We also consider a simple empirical weight matrix estimator based on an exchangeable approximation to the covariance structure. In asymptotic relative efficiency calculations and simulations, it is found that for the scenarios examined, the gains in efficiency using the optimal weights are typically quite small relative to the other weights, suggesting that it may not be useful in practice to attempt to estimate the optimal weights. The simple empirical weights are also found to have good efficiency, and thus provide a practically useful estimator of β.

In the next section the piecewise constant family of weights is defined and the optimal weights in the family derived. A simpler version based on an 'exact' martingale equation is also given, and found to be equivalent to the optimal weights at $\beta = 0$. Constant weights are discussed in Section 3, the empirical weight estimator is given in Section 4, and formulas for specific models are given in Section 5. Numerical results are given in Section 6.

2. Optimal Weights

For a simpler presentation, we hereafter consider the case where z_{ij} and β are scalar and the cluster size is constant ($n_i \equiv n$). Most results can be extended to higher dimensional covariates and non-constant n_i, though the computations become substantially more complicated. Define $a_{ij}(t, \beta) = \Sigma_k z_{ik}\, \omega_{ikj}(t, \beta)$, and restrict the integral to a finite time interval $[0, T]$, where T is a fixed constant, giving the equations

$$D(\beta) = \sum_i \sum_j \int_0^T a_{ij}(t, \beta) d\hat{M}_{ij}(t, \beta) = 0. \quad (3)$$

In this formulation, the problem of choosing the unknown weight functions is equivalent to choosing the functions $a_{ij}(t, \beta)$. We restrict attention to functions $a_{ij}(t, \beta)$ which are constant on prespecified time intervals $[T_r, T_{r+1})$, with $T_1 = 0 < T_2 < \cdots < T_{L+1} = T$. That is, $a_{ij}(t, \beta) = \Sigma_{r=1}^{L} a_{ij}^r I(T_r \leq t < T_{r+1})$ for scalars $a_{ij}^r = a_{ij}^r(\beta)$.

We assume that the true β is in the interior of a compact set $\mathcal{B} \subset R^1$, that the a_{ij}^r are bounded and differentiable on $\mathcal{B}$, and that the covariates z_{ij} are bounded. The main objective of this paper is to to calculate the values of the a_{ij}^r that minimize the asymptotic variance of the solution $\hat{\beta}$ of (3), and to study the performance of other weights relative to these optimal weights. The asymptotic framework considered is as $N \to \infty$ with n held fixed.

2.1. Asymptotic Variance

Define $\bar{a}(t, \beta) = \Sigma_{i=1}^{N} \Sigma_{j=1}^{n} \pi_{ij}(t) e^{z_{ij}\beta} a_{ij}(t, \beta) / \Sigma_{i=1}^{N} \Sigma_{j=1}^{n} \pi_{ij}(t) e^{z_{ij}\beta}$, where $\pi_{ij}(t) = P(X_{ij} \geq t) = \mathrm{E}\{Y_{ij}(t)\}$, and set $D_i = \Sigma_{j=1}^{n} \int_0^t \{a_{ij}(t, \beta) - \bar{a}(t, \beta)\}\, dM_{ij}(t, \beta)$. Under the assumptions listed earlier and the regularity assumptions outlined in Cai and Prentice (1997),

$N^{-1/2}|D(\beta)-\Sigma_i D_i| \to 0$ in probability, so $D(\beta)$ and $\Sigma_i D_i$ are asymptotically equivalent, and the asymptotic variance of $\sqrt{N}(\hat{\beta}-\beta)$ is

$$\lim_{N\to\infty} NA^{-1}\left\{\sum_{i=1}^{N}\mathrm{Var}(D_i)\right\}A^{-1}, \tag{4}$$

where

$$A=\sum_{i,j}\int_0^T\{a_{ij}(t,\beta)-\bar{a}(t,\beta)\}z_{ij}\pi_{ij}(t)e^{zij\beta}d\Lambda_0(t). \tag{5}$$

Since the $a_{ij}(t, \beta)-\bar{a}(t, \beta)$ are fixed functions, it is straightforward to show that

$$\mathrm{Var}(D_i)=\sum_{j,k}\int_0^T\int_0^T\{a_{ij}(u,\beta)-\bar{a}(u,\beta)\}\{a_{ik}(v,\beta)-\bar{a}(v,\beta)\}C_{ijk}(du,dv), \tag{6}$$

where $C_{ijk}(u, v) = \mathrm{Cov}\{M_{ij}(u, \beta), M_{ik}(v, \beta)\}$.

For H from (1), define

$$\lambda^{rl}(u,v)=\frac{(-1)^{r+l}}{H(u,v)}\frac{\partial^{r+l}H(u,v)}{\partial u^r\partial v^l},\ r,l=0,1. \tag{7}$$

From standard martingale results (eg Theorem 2.5.3 of Fleming and Harrington, 1991),

$$C_{ijj}(s,t)\equiv E\{M_{ij}(s,\beta)M_{ij}(t,\beta)\}=\int_0^{s\wedge t}e^{z_{ij}\beta}\pi_{ij}(u)\lambda_0(u)du,$$

and it is shown in the Appendix A that

$$C_{ijk}(s,t)\equiv E\{M_{ij}(s,\beta)M_{ik}(t,\beta)\}=\int_0^s\int_0^t\pi_{ijk}(u,v)c_{ijk}(u,v)d\Lambda_0(u)d\Lambda_0(v) \tag{8}$$

for $j \neq k$, where

$$\begin{aligned}c_{ijk}(u,v)=[&\lambda^{11}\{\Lambda_{ij}(u),\Lambda_{ik}(v)\}-\lambda^{10}\{\Lambda_{ij}(u),\Lambda_{ik}(v)\}\\&-\lambda^{01}\{\Lambda_{ij}(u),\Lambda_{ik}(v)\}+1]e^{z_{ij}\beta+z_{ik}\beta},\end{aligned}$$

$\pi_{ijk}(u, v) = P(X_{ij} > u, X_{ik} > v)$, and $\Lambda_{il}(w) = \Lambda_0(w)e^{z_{ij}\beta}$. The function $c_{ijk}(u, v)$ is the conditional martingale covariance rate function of Prentice and Cai (1992). If failure times within a cluster are independent, then $c_{ijk}(u, v) = 0$.

2.2. Time-Dependent Optimal Weights

To get tractable formulas, we approximate $\bar{a}(t, \beta) \doteq \bar{a}^r \equiv \bar{a}\{(T_r + T_{r+1})/2, \beta\}$ for $T_r \leq t < T_{r+1}$ throughout the following. When the lengths of the intervals are sufficiently small, the effect of this approximation will be small. Also, define $\tilde{a}_{ij}^r = a_{ij}^r - \bar{a}_r$, $C_{ijj}^r = C_{ijj}(T_{r+1}, T_{r+1}) - C_{ijj}(T_r, T_r)$,

$$C_{ijk}^{r_1, r_2} = \int_{T_{r_1}}^{T_{r_1+1}} \int_{T_{r_2}}^{T_{r_2+1}} \pi_{ijk}(u, v) c_{ijk}(u, v) d\Lambda_0(u) d\Lambda_0(v)$$

for $j \neq k$, and $\tilde{z}_{ij}^r = z_{ij} C_{ijj}^r$. In the numeric calculations later, the C_{ijj} and C_{ijk} are evaluated using Gaussian quadrature. In addition, let $\tilde{a}_{ik} = (\tilde{a}_{ik}^1, \ldots, \tilde{a}_{ik}^L)$, $\tilde{a}_i = (\tilde{a}_{i1}', \ldots, \tilde{a}_{in}')'$ and $\tilde{a} = (\tilde{a}_1', \ldots, \tilde{a}_N')'$, and similarly define $\tilde{z}_{ik}$, $\tilde{z}_i$ and $\tilde{z}$ in terms of the $\tilde{z}_{ik}^r$. Then from (6),

$$\begin{aligned} \mathrm{Var}(D_i) &= \sum_{j=1}^{n} \sum_{r=1}^{L} (a_{ij}^r - \bar{a}^r)^2 C_{ijj}^r + \sum_{k \neq j} \sum_{r_1, r_2} (a_{ij}^{r_1} - \bar{a}^{r_1})(a_{ik}^{r_2} - \bar{a}^{r_2}) C_{ijk}^{r_1, r_2} \\ &= \sum_{j,k} \tilde{a}_{ij}' Q_{ijk} \tilde{a}_{ik} = \tilde{a}_i' Q_i \tilde{a}_i, \end{aligned}$$

where $Q_{ijj} = \mathrm{diag}(C_{ijj}^1, \ldots, C_{ijj}^L)$, $Q_{ijk} = (C_{ijk}^{r_1, r_2})_{L \times L}$ for $j \neq k$, and

$$Q_i = \begin{pmatrix} Q_{i11} & \cdots & Q_{i1n} \\ \vdots & \vdots & \vdots \\ Q_{in1} & \cdots & Q_{inn} \end{pmatrix},$$

and from (5), $A = \Sigma_{i,j} \Sigma_r \tilde{a}_{ij}^r \tilde{z}_{ij}^r = \tilde{a}'\tilde{z}$. Note that each Q_i is symmetric and positive definite, subject to mild regularity conditions on the composition of the clusters. Hence, the asymptotic variance (4) can be approximated by

$$N\tilde{a}'Q\tilde{a}/(\tilde{a}'\tilde{z})^2, \tag{9}$$

where $Q = \mathrm{diag}(Q_1, \ldots, Q_N)$. We only consider this expression for finite N. This can either be thought of as an approximation to the limiting variance when clusters are sampled from an infinite population of possible clusters, or as the limiting variance when the population contains only a finite number of possible cluster types in the same proportions as in (9).

Since the elements of $\tilde{a}$ have the form $a_{ij}^r - \bar{a}_r$, $\tilde{a}$ must satisfy the constraints $\Sigma_{ij} \tilde{a}_{ij}^r \pi_{ij}^r e^{z_{ij}\beta} = W_r'\tilde{a} = 0$, $r = 1\ldots, L$, where $\pi_{ij}^r = \pi_{ij}\{(T_r + T_{r+1})/2\}$ and the $[r + \{(i - 1)n + j - 1\}L]$th element of W_r is $\pi_{ij}^r e^{z_{ij}\beta}$, $j = 1, \ldots, n$, $i = 1, \ldots, N$, with the other elements = 0. Since the scaling of $\tilde{a}$ is arbitrary, the piecewise constant optimal weights can thus be found by first finding the solution to

$$\min_{\tilde{a}} \ \tilde{a}'Q\tilde{a}, \tag{10}$$

subject to $\tilde{a}'\tilde{z} = 1$ and $W_r'\tilde{a} = 0$, $r = 1,\ldots, L$. Then any set of a_{ij}^r giving this $\tilde{a}$ will be an optimal set of piecewise constant weights. In particular, we can take $a_{ij}^r = \tilde{a}_{ij}^r$.

Solving the constrained optimization problem (10) is equivalent to solving the linear system

$$\begin{aligned} 2Q\tilde{a} + \sum_{r=1}^{L} \psi_r W_r + \psi_{L+1}\tilde{z} &= 0 \\ W_r'\tilde{a} &= 0, \ r = 1, \ldots, L \\ \tilde{a}'\tilde{z} &= 1, \end{aligned}$$

where the ψ_r, $r = 1,\ldots, L + 1$, are Lagrange multipliers. The values of $\psi = (\psi_1,\ldots, \psi_{L+1})'$ can be obtained from $F'Q^{-1}F\psi = -2b$, where $F = (W_1,\ldots, W_L, \tilde{z})$ and $b = (0,\ldots, 0, 1)'$. Then $\tilde{a} = -\frac{1}{2}Q^{-1}F\psi$. The estimator computed from these weights will be denoted $\hat{\beta}_{Opt}$.

When the observations are independent, $c_{ijk}(u, v) = 0$ for $j \neq k$, and the Q_i are diagonal matrices. In this case, it is easily verified from (9) that the optimal weight functions are $a_{ij}(t, \beta) \propto z_{ij}$, giving the usual partial likelihood scores for independent data. This is not surprising, since it is well-known that the partial likelihood is semiparametric efficient for independent data. The estimator computed from the independence working model will be denoted $\hat{\beta}_{Ind}$.

Simpler weights can be obtained by considering the ‘exact’ martingale estimating equation

$$0 = D_0(\beta) = \sum_{i=1}^{N} \int_0^T \sum_{j=1}^{n} a_{ij}(t, \beta) dM_{ij}(t, \beta). \tag{11}$$

This is not a practically useful equation because of the dependence of the M_{ij} on Λ_0. It is easily verified that for this equation the asymptotic variance of $N^{1/2}(\hat{\beta}-\beta)$ is lim $N\mathrm{Var}\{D_0(\beta)\}/\mathrm{E}\{\nabla\, D_0(\beta)\}^2$, and that

$$E\{\nabla D_0(\beta)\} = \sum_{ij} \int_0^T a_{ij}(t, \beta) z_{ij} \pi_{ij}(t) e^{z_{ij}\beta} d\Lambda_0(t).$$

Again using the step function approximations given above, it follows that the asymptotic variance is $\alpha'Q\alpha/(\alpha'\tilde{z})^2$, where the vector α has components a_{ij}^r. From the Cauchy-Schwarz inequality, $\alpha'Q\alpha/(\alpha'\tilde{z})^2 \geq 1/(\tilde{z}'Q^{-1}\tilde{z})$, and since this bound is attained when $\alpha = Q^{-1}\tilde{z}$, it follows that these are the optimal weights for (11). These weights can also be used in the estimated martingale equation, and the solution to (3) using these weights will be denoted $\hat{\beta}_{EM}$. It is shown in Appendix B that at $\beta = 0$, and under equal censoring (as assumed throughout), the centered weights $\tilde{a}$ obtained from $\alpha = Q^{-1}\tilde{z}$ are also a solution to (10). It will also be seen in the numerical results later that $\hat{\beta}_{EM}$ is highly efficient for nonzero β, too.

3. Time Independent Weights

In this section, we consider the time independent weights in the estimating equation. That is, $a_{ij}(t, \beta) \equiv a_{ij}(\beta)$, or equivalently $a_{ij}^1 = \ldots = a_{ij}^L = a_{ij}^*$. We continue to use the piecewise constant approximation for $\bar{a}(t, \beta)$ defined in the previous section.

Let $a^* = (a_{11}^*, a_{12}^*, \ldots, a_{N_n}^*)'$ and $U_r = (\Sigma_{ij}\pi_{ij}^r e^{z_{ij}\beta})^{-1}(\pi_{11}^r e^{z_{11}\beta}, \pi_{12}^r e^{z_{12}\beta}, \ldots, \pi_{Nn}^r e^{z_{Nn}\beta})'$, and note $\bar{a}_r = U_r' a^*$. Also, let $\mathbf{1}_l$ denote a length l vector of 1's, and I_l the $l \times l$ identity matrix. Then

$$\tilde{a}_{ij} = \left(a_{ij}^* - \bar{a}^1, \ldots, a_{ij}^* - \bar{a}^L\right)' = a_{ij}^* \otimes \mathbf{1}_L - U'a^*,$$

where $U = (U_1, \ldots, U_L)$ and $\otimes$ denotes the tensor product, and $\tilde{a} = a^* \otimes \mathbf{1}_L - \mathbf{1}_{Nn} \otimes U'a^* = Pa^*$, where $P = I_{Nn} \otimes \mathbf{1}_L - \mathbf{1}_{Nn} \otimes U'$. Thus in this setting we can express (9) as $N(a^*)'P'QPa^*/(\tilde{z}'Pa^*)^2$, and the optimization problem is to find a^* minimizing $(a^*)'P'QPa^*$ subject to $\tilde{z}'Pa^* = c$.

Equivalently, we need to solve for a^* in

$$\begin{aligned} 2P'QPa^* + \psi P'\tilde{z} &= 0 \\ \tilde{z}'Pa^* &= c, \end{aligned} \tag{12}$$

where again ψ is a Lagrange multiplier. The rank of P is $Nn - 1$, so $P'QP$ is singular. Since Q is positive definite, the range space of $P'QP$ is the same as that of P', so solutions exist to the first of these equations, and a particular solution is $a^* = (-\psi/2)(P'QP)^- P'\tilde{z}$, where $(P'QP)^-$ is the Moore-Penrose generalized inverse of $P'QP$ (see Rao and Mitra, 1971). Since the scaling of a^* is arbitrary, the value of $c \neq 0$, and hence of ψ, is arbitrary, and we set $\psi = 1$. The estimator defined from these weights will be denoted $\hat{\beta}_{OC}$.

For the exact martingale equation (11), the optimal constant weights are given by

$$a_i^* = V_i^{-1} z_i^*, \quad i = 1, \ldots, N, \tag{13}$$

where $a_i^* = (a_{i1}^*, \ldots, a_{in}^*)'$, $z_i^* = (z_{i1}^*, \ldots, z_{in}^*)'$, $z_{ij}^* = z_{ij}C_{ijj}(T, T)$, and V_i has j, j' component $\mathrm{Cov}\{M_{ij}(T, \beta), M_{ij'}(T, \beta)\}$. Since V_i is a covariance matrix, and $z_i^* = \mathrm{E}\{\partial M_i(T, \beta)/\partial\beta\}$, these are similar in form to the optimal weights in standard generalized estimating equations (Liang and Zeger, 1986). It can again be shown that (13) is equivalent to the solution to (12) at $\beta = 0$, and thus are optimal constant weights in that case. The estimator defined from these weights will be denoted $\hat{\beta}_{EMC}$.

4. Empirical Weights

In this section, empirical weights that could be used in practice are given. We base these on the constant exact martingale weights (13), but further approximate V_i with an equal variance exchangeable covariance structure:

$$V = \sigma^2\left\{(1-\rho)I_n + \rho\mathbf{1}_n\mathbf{1}_n'\right\}.$$

Since the variances and covariances do depend on the covariates, unless $\beta = 0$, this is a further approximation, but reduces the number of parameters to be estimated. The quantities $M_{ij}(t)$ and z_{ij}^*, defined in the previous section, can be estimated by

$$\hat{M}_{ij}(t) = N_{ij}(t) - e^{z_{ij}\hat{\beta}}\hat{\Lambda}_0(X_{ij} \wedge t)$$

and

$$\hat{z}_{ij}^* = z_{ij}\int_0^T \hat{\pi}_{ij}(t)e^{z_{ij}\hat{\beta}}d\hat{\Lambda}_0(t),$$

where $\hat{\pi}_{ij}(t) = \exp\{-\hat{\Lambda}_0(t)e^{z_{ij}\hat{\beta}}\}\hat{G}(t)$, $\hat{\beta}$ and $\hat{\Lambda}_0(\cdot)$ are the standard estimators for β and Λ_0 obtained from fitting the marginal proportional hazards model, and $\hat{G}(t) = \exp\{-\hat{\Lambda}_c(t)\}$, with $\hat{\Lambda}_c$ the Nelson estimator of the cumulative hazard of the censoring distribution based on pairs $(X_{ij}, 1 - \delta_{ij})$. Then σ^2 in V is estimated by $\Sigma_{ij}\ \hat{M}_{ij}(T)^2/(Nn)$, and the covariance $\rho\sigma^2$ by

$$\sum_i \sum_{j<k} \hat{M}_{ij}(T)\hat{M}_{ik}(T)\{Nn(n-1)/2\}.$$

Finally, we apply (13), with V_i and z_i^* replaced by these estimates, to obtain the empirical weights $\hat{a}_i^*$. The estimator obtained from these weights is denoted by $\hat{\beta}_{Emp}$. We note that these weights are consistent for the optimal constant weights when $\beta = 0$.

5. Specific Survival Models

Specifying a joint distribution for clustered data is most easily accomplished using an Archimedean copula model. We consider two special cases, the Clayton and positive stable models.

In Clayton's (1978) model (see also Oakes, 1982, and Clayton and Cuzick, 1985), the function H in (1) is $H(u, w) = b(u, w)^{-1/\theta}$, where $b(u, w) = e^{\theta u} + e^{\theta w} - 1$. Then from (7), $\lambda^{10}(u, w) = e^{\theta u}/b(u, w)$, $\lambda^{01}(u, w) = e^{\theta w}/b(u, w)$ and $\lambda^{11}(u, w) = (1 + \theta)\, e^{\theta u}e^{\theta w}/b(u, w)^2$.

Let $S(t|z) = \exp\{-\Lambda_0(t)e^{z\beta}\}$ be the marginal survivor function. Defining $B_{ijk}(u, w) = S(u|z_{ij})^{-\theta} + S(w|z_{ik})^{-\theta} - 1$, the joint survival function from Clayton's model is $S_{ijk}(u, w) = B_{ijk}(u, w)^{-1/\theta}$. Also, $\pi_{ijk}(u, w) = S_{ijk}(u, w)G(u)G(w)$. Then from (8), the martingale covariance function can be written

$$
\begin{aligned}
C_{ijk}(s,t) = & \int_0^s \int_0^t S_{ijk}(u,w)G(u)G(w) \\
& \times \left\{ (1+\theta)\frac{S(u|z_{ij})^{-\theta}S(w|z_{ik})^{-\theta}}{B_{ijk}(u,w)^2} - \frac{S(u|z_{ij})^{-\theta}}{B_{ijk}(u,w)} - \frac{S(w|z_{ik})^{-\theta}}{B_{ijk}(u,w)} + 1 \right\} \\
& \times e^{z_{ij}\beta + z_{ik}\beta} d\Lambda_0(w) d\Lambda_0(u).
\end{aligned}
$$

Kendall's τ, defined as the probability that the components of the difference of independent bivariate pairs $(T_{11}, T_{12}) - (T_{21}, T_{22})$ have the same sign minus the probability that they have opposite signs, is equal to $\theta/(\theta + 2)$ for this model. The natural multivariate extension, which has the bivariate marginals given above, is given by

$$
P(T_{i1} > t_1, \ldots, T_{in} > t_n | z_{i1}, \ldots, z_{in}) = \left\{ \sum_{j=1}^{n} S(t_j|z_{ij})^{-\theta} - n + 1 \right\}^{-1/\theta} \tag{14}
$$

For the positive stable model, $H(u, w) = \exp\{-(u^{1/\alpha} + w^{1/\alpha})^{\alpha}\}$, where $0 < \alpha < 1$. Smaller values of α indicate stronger association, and $\alpha = 1$ yields independence. Also, Kendall's τ for this model is $1 - \alpha$. Here $\lambda^{10}(u, w) = (u^{1/\alpha} + w^{1/\alpha})^{\alpha-1}u^{1/\alpha-1}$, $\lambda^{01}(u, w) = (u^{1/\alpha} + w^{1/\alpha})^{\alpha-1}w^{1/\alpha-1}$, and

$$
\lambda^{11}(u,w) = u^{1/\alpha-1}w^{1/\alpha-1}(u^{1/\alpha} + w^{1/\alpha})^{\alpha-2}\{(u^{1/\alpha} + w^{1/\alpha})^{\alpha} - (\alpha - 1)/\alpha\}.
$$

The martingale covariance function can again be obtained by substituting these expressions in (8), as above.

6. Numerical Results

We first compute asymptotic relative efficiencies $\text{ARE}(\hat{\beta}, \hat{\beta}_{Opt}) = \text{Var}(\hat{\beta}_{Opt}) / \text{Var}(\hat{\beta})$ for the estimators $\hat{\beta}$ using various weight functions. It is only possible to compute the optimal weights with a finite number of possible cluster configurations. We used a population of 80 clusters, and generated random $U(0, 1)$ covariate values within each cluster. Each of these 80 possible clusters is then assumed to occur with equal probability as $N \rightarrow \infty$. The calculations were repeated for ten different populations of covariate values for each

Table 1. Asymptotic efficiencies of $\hat{\beta}_{OC}$ (optimal constant) and $\hat{\beta}_{Ind}$ (independence weights) relative to $\hat{\beta}_{Opt}$ (optimal weights); $\beta = 1$. Entries are the range of the AREs over 10 random covariate configurations.

	$n = 2$		$n = 3$	
Model	ARE($\hat{\beta}_{OC}$; $\hat{\beta}_{Opt}$)	ARE($\hat{\beta}_{Ind}$; $\hat{\beta}_{Opt}$)	ARE($\hat{\beta}_{OC}$; $\hat{\beta}_{Opt}$)	ARE($\hat{\beta}_{Ind}$; $\hat{\beta}_{Opt}$)
Clayton $\theta = 1$	(1.00,1.00)	(.86,.86)	(1.00,1.00)	(.79,.82)
Clayton $\theta = 2$	(.99,.99)	(.73,.74)	(.99,.99)	(.64,.68)
Clayton $\theta = 3$	(.97,.97)	(.64,.65)	(.96,.97)	(.56,.60)
Clayton $\theta = 4$	(.95,.95)	(.59,.60)	(.94,.94)	(.50,.55)
Stable $\alpha = .9$	(1.00,1.00)	(.99,.99)	(.99,.99)	(.97,.98)
Stable $\alpha = .6$	(.93,.94)	(.79,.81)	(.92,.93)	(.72,.76)
Stable $\alpha = .3$	(.77,.78)	(.49,.50)	(.78,.79)	(.43,.47)

scenario. We used $\Lambda_0(u) = u$ and exponential censoring with $G(u) = e^{-u}$. We also truncate follow-up at $T = 3$ to restrict attention to a finite interval. Cluster sizes of $n = 2$ and 3 were considered.

In the calculations, we set $L = 50$ and $T_r = (r - 1)T/L$, $r = 1, \ldots, L + 1$. The C^{r}_{ijj} and C^{r_1,r_2}_{ijk} were evaluated using 4 point Gauss-Legendre quadrature, except for the positive stable model with smaller values of α, where a 25 point rule was used in the first interval because the integrand has substantial mass along a narrow ridge there.

Table 1 gives results for $\beta = 1$. The ARE's show that $\hat{\beta}_{Ind}$ can lose substantial efficiency as the degree of dependence increases, and even in the Clayton model with $\theta = 1$ (Kendall's $\tau = 1/3$), $\hat{\beta}_{Ind}$ is only 80% as efficient as the optimal and optimal constant weights. Except under very strong dependence, ARE($\hat{\beta}_{OC}$, $\hat{\beta}_{Opt}$) $> .9$, suggesting that typically there is little to be gained by considering nonconstant weights. The asymptotic variances were also computed for $\hat{\beta}_{EM}$ and $\hat{\beta}_{EMC}$. In all scenarios considered, ARE($\hat{\beta}_{EM}$, $\hat{\beta}_{Opt}$) $\geq .99$ and ARE($\hat{\beta}_{EMC}$, $\hat{\beta}_{OC}$) $\geq .99$, suggesting that the simpler weights obtained from the exact martingale equations should be adequate in general. Similar calculations were also done for $\beta = 0$. The patterns were generally similar, with slightly higher ARE's for both $\hat{\beta}_{EM}$ and $\hat{\beta}_{IND}$ in the Clayton model and slightly lower ARE's for both in the stable model. As noted above, at $\beta = 0$, ARE($\hat{\beta}_{EM}$, $\hat{\beta}_{Opt}$) = ARE($\hat{\beta}_{EMC}$, $\hat{\beta}_{OC}$) = 1.00.

We next used simulations to examine finite sample performance of the estimators for different weight functions. Data was generated from the models specified above, with $N = 80$ clusters used in each scenario, and the same iid $U(0, 1)$ set of covariate values used throughout for each value of n. The estimators $\hat{\beta}_{Opt}$, $\hat{\beta}_{EM}$, $\hat{\beta}_{EMC}$, $\hat{\beta}_{OC}$, $\hat{\beta}_{Emp}$, and $\hat{\beta}_{Ind}$ were computed for each sample, and their variances estimated with the empirical variances of the estimators computed from 10,000 simulated samples. The results are given in Table 2 for $\beta = 1$. The variances for $\hat{\beta}_{OC}$ are always very close to those for $\hat{\beta}_{EMC}$, and are omitted. The results indicate that the efficiency gains suggested by the ARE's can largely be obtained with moderate size samples, although the gains are slightly smaller with this sample size. The performance of $\hat{\beta}_{Emp}$ is nearly as good as $\hat{\beta}_{EMC}$, suggesting that it would often be adequate in practice. Results for $\beta = 0$ (omitted) were generally similar, again with slightly higher efficiencies for $\hat{\beta}_{EMC}$, $\hat{\beta}_{Emp}$ and $\hat{\beta}_{Ind}$

Table 2. Var($\hat{\beta}$) and ratios Var($\hat{\beta}_{Opt}$) / Var($\hat{\beta}$) for estimators $\hat{\beta}$ obtained from various weight functions; $N = 80$, $\beta = 1$. Standard errors of the variances estimates are about .002 for $n = 2$ and. 001 for $n = 3$.

			$\hat{\beta}_{EM}$		$\hat{\beta}_{EMC}$		$\hat{\beta}_{Emp}$		$\hat{\beta}_{Ind}$	
n	Model	Var($\hat{\beta}_{Opt}$)	Var	Ratio	Var	Ratio	Var	Ratio	Var	Ratio
2	Clayton $\theta = 0$	.128	.128	1.00	.128	1.00	.130	.98	.128	1.00
	Clayton $\theta = 1$	.110	.111	.99	.111	.99	.112	.98	.126	.87
	Clayton $\theta = 2$	.094	.095	.99	.096	.98	.097	.97	.127	.74
	Clayton $\theta = 3$	.084	.085	.99	.087	.97	.088	.96	.128	.66
	Clayton $\theta = 4$	.078	.078	.99	.082	.95	.083	.94	.128	.60
3	Clayton $\theta = 0$	.086	.086	1.00	.086	1.00	.087	.99	.086	1.00
	Clayton $\theta = 1$	.069	.069	.99	.069	.99	.070	.98	.085	.81
	Clayton $\theta = 2$	.057	.057	.99	.058	.98	.059	.97	.085	.67
	Clayton $\theta = 3$	.050	.050	.99	.052	.96	.052	.95	.086	.58
	Clayton $\theta = 4$	.046	.046	.99	.049	.93	.050	.93	.087	.53
2	Stable $\alpha = 1$	.128	.128	1.00	.128	1.00	.130	.98	.128	1.00
	Stable $\alpha = .9$	.126	.126	1.00	.126	.99	.128	.98	.128	.98
	Stable $\alpha = .6$	.103	.103	1.00	.110	.93	.113	.91	.131	.79
	Stable $\alpha = .3$	.067	.068	.99	.086	.79	.087	.78	.134	.50
3	Stable $\alpha = 1$	.086	.086	1.00	.086	1.00	.087	.99	.086	1.00
	Stable $\alpha = .9$	.081	.081	1.00	.082	.99	.083	.98	.083	.98
	Stable $\alpha = .6$	.062	.062	1.00	.067	.93	.067	.92	.082	.75
	Stable $\alpha = .3$	.039	.039	.99	.049	.80	.049	.79	.084	.46

relative to the optimal weights in the Clayton model and slightly lower efficiencies in the stable model.

7. Discussion

In this paper we have considered the choice of weight function in estimating equations for marginal proportional hazards models for clustered data. The optimal and optimal constant weights and the corresponding weights derived from the exact martingale equation require knowledge of the true distributions, and are thus not practically useful. The main purpose of this investigation was to see what gains in efficiency might be possible if the optimal weights were known, and also to gain some insight into what types of weight functions we should be trying to estimate.

In ARE calculations, we found that the gains in efficiency of the weighted equations relative to the standard partial likelihood scores can be substantial with moderate dependence. This is consistent with the results of other investigations, such as those of Cai and Prentice (1995, 1997). The differences between $\hat{\beta}_{Opt}$ and $\hat{\beta}_{EM}$ and between $\hat{\beta}_{OC}$ and $\hat{\beta}_{EMC}$ were very small in all scenarios examined. The estimators based on constant weights performed well except in the positive stable model with strong dependence. A simple empirical weight estimator also performed nearly as well as $\hat{\beta}_{OC}$. This suggests that in many settings there may be little advantage to attempting to estimate the more complex optimal weights. While the particular form of the estimated weights used here performed well in our simulations, we suspect that other approaches, such as those of Cai and Prentice (1995, 1997), will also often perform well in practice.

Our results have several limitations. We assumed that covariates affect the joint distributions only through the marginal proportional hazards model, and assumed exchangeable symmetry of generalized residuals within a cluster. The effect of more complex forms of dependence is unknown, but it would likely be even more difficult to calculate and estimate optimal weights under more complex models. The distribution of covariates used in the calculations was also independent of cluster membership. Generally the covariate configuration will affect the efficiency of the weighted estimating equations, with higher efficiency with large intra-cluster covariate differences and lower efficiency when covariate values have positive intra-cluster correlation. Finally, throughout only fixed covariates were considered. The theoretical results on optimal weights in Section 2 are easily extended to time-varying covariates that are approximately constant on short time intervals, but this would require specification of full covariate trajectories for the computations. The consideration of constant weights and the form of the empirical estimator do not generalize to time-varying covariates, though.

Acknowledgments

This work was supported by grant CA57253, awarded by the National Cancer Institute, USA.

Appendix A. Derivation of $C_{ijk}(u, v)$

For $j \neq k$,

$$E\{M_{ij}(s)M_{ik}(t)\} = P(X_{ij} \leq s, X_{ik} \leq t, \delta_{ij} = 1, \delta_{ik} = 1)$$
$$- \int_0^s P(X_{ij} \geq u, X_{ik} \leq t, \delta_{ik} = 1)e^{z_{ij}\beta} d\Lambda_0(u)$$
$$- \int_0^t P(X_{ij} \leq s, X_{ik} \geq w, \delta_{ij} = 1)e^{z_{ik}\beta} d\Lambda_0(w)$$
$$+ \int_0^s \int_0^t P(X_{ij} \geq u, X_{ik} \geq w)e^{z_{ij}\beta + z_{ik}\beta} d\Lambda_0(w) d\Lambda_0(u). \tag{15}$$

Now $P(X_{ij} \geq u, X_{ik} \geq w) = \pi_{ijk}(u, w)$. Using (1) and (7),

$$P(X_{ij} \leq s, X_{ik} \leq t, \delta_{ij} = 1, \delta_{ik} = 1) = \int_0^s \int_0^t G(u)G(w)S_{ijk}(du, dw)$$
$$= \int_0^s \int_0^t \pi_{ijk}(u, w)\lambda^{11}\{\Lambda_0(u)e^{z_{ij}\beta}, \Lambda_0(w)e^{z_{ik}\beta}\}e^{z_{ij}\beta + z_{ik}\beta} d\Lambda_0(u) d\Lambda_0(w).$$

Similarly,

$$P(X_{ij} \geq u, X_{ik} \leq t, \delta_{ik} = 1) = \int_0^t G(u)G(w)S_{ijk}(u, dw)$$
$$= \int_0^t \pi_{ijk}(u, w)\lambda^{01}\{\Lambda_0(u)e^{z_{ij}\beta}, \Lambda_0(w)e^{z_{ik}\beta}\}e^{z_{ik}\beta} d\Lambda_0(w),$$

and

$$P(X_{ij} \leq s, X_{ik} \geq w, \delta_{ij} = 1) = \int_0^s G(u)G(w)S_{ijk}(du, w)$$
$$= \int_0^t \pi_{ijk}(u, w)\lambda^{10}\{\Lambda_0(u)e^{z_{ij}\beta}, \Lambda_0(w)e^{z_{ik}\beta}\}e^{z_{ij}\beta} d\Lambda_0(u).$$

Substituting these expressions in (15) then gives (8). This can also be obtained from formula (2) of Prentice and Cai (1992).

Appendix B. Optimality of $Q^{-1}\tilde{z}$ at $\beta = 0$.

Define $\bar{z}(t, \beta) = \Sigma_{i=1}^{N}\Sigma_{j=1}^{n}\pi_{ij}(t)z_{ij}\, e^{z_{ij}\beta} / \Sigma_{i=1}^{N}\Sigma_{j=1}^{n}\pi_{ij}(t)e^{z_{ij}\beta}$ and $z_{ij}^{r} = \bar{z}\{(T_r + T_{r+1})/2, \beta\}$, and let ζ be the vector with components $\zeta_{ij}^{r} = (z_{ij} - \bar{z}^{r})C_{ijj}^{r}$. Since (5) equals

$$\sum_{i,j}\int_0^T \{a_{ij}(t,\beta) - \bar{a}(t,\beta)\}\{z_{ij} - \bar{z}(t,\beta)\}\pi_{ij}(t)e^{z_{ij}\beta}d\Lambda_0(t),$$

the piecewise constant approximation also leads to the formula

$$\tilde{a}'Q\tilde{a}/(\tilde{a}'\zeta)^2 \tag{16}$$

for the asymptotic variance. In this appendix it is shown that if $\beta = 0$, and under equal censoring, $\tilde{a}$ obtained from $\alpha = Q^{-1}\tilde{z}$ minimizes (16) subject to the constraints $W_r'\tilde{a} = 0$. This is done by showing first that $\tilde{a} = Q^{-1}\zeta$ minimizes (16) and satisfies the constraints, and then that $\alpha = Q^{-1}\tilde{z}$ gives $\tilde{a} = Q^{-1}\zeta$.

From the Cauchy-Schwarz inequality, the global minimum of (16) is attained by $\tilde{a} = Q^{-1}\zeta$. To show that this value also satisfies the constraints, first, when $\beta = 0$, and under equal censoring, π_{ij}, C_{ijj}^{r} and $C_{ijk}^{r_1,r_2}$ do not depend on i, j, k, so $\bar{a}_r = \Sigma_{i,j}\, a_{ij}^{r}/(Nn)$ and $\bar{z}^{r} = \Sigma_{i,j}z_{ij}/(Nn)$, and the Q_{ijj} and $Q_{ijk}, j \neq k$, do not depend on i, j, k. Thus $\Sigma_{ij}\,\tilde{a}_{ij} = 0$ and $\Sigma_{ij}\,\zeta_{ij}^{r} = \Sigma_{ij}\,(z_{ij} - \bar{z}^{r})\,C_{ijj}^{r} = 0$. Also, the nonzero components of W_r are all equal to some constant c. Since the $Q_{ijj'}$ matrices do not depend on i,j, the matrices Q_i^{-1} are all equal, and each is of the form

$$\begin{pmatrix} E & B & \cdots & B \\ B & E & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & E \end{pmatrix},$$

where E and B are $L \times L$ matrices. Thus setting $\zeta_{ij} = (\zeta_{ij}^{1}, \ldots, \zeta_{ij}^{L})'$,

$$Q^{-1}\zeta = \begin{pmatrix} (E-B)\zeta_{11} + B\sum_{j'}\zeta_{1j'} \\ \vdots \\ (E-B)\zeta_{Nn} + B\sum_{j'}\zeta_{Nj'} \end{pmatrix}.$$

Letting E_r' and B_r' be the rth rows of E and B,

$$\begin{aligned} W_r' Q^{-1}\zeta &= c\left(\sum_{i,j}\left(E_r' - B_r'\right)\zeta_{ij} + \sum_{i,j,j'} B_r'\zeta_{ij'}\right) \\ &= c\left(E_r' - B_r'\right)\sum_{i,j}\zeta_{ij} + ncB_r'\sum_{i,j'}\zeta_{ij'} \\ &= 0. \end{aligned} \tag{17}$$

Thus $\tilde{a} = Q^{-1}\zeta$ satisfies $W_r{}'\tilde{a} = 0$, $r = 1, \ldots, L$, and thus minimizes (16) subject to these constraints.

Now consider $\alpha = Q^{-1}\tilde{z}$. Similar to (17), for these a_{ij}^r,

$$\begin{aligned} a_{ij}^r - \bar{a}^r &= a_{ij}^r - \tfrac{1}{Nn}\sum_{hk} a_{hk}^r \\ &= (E_r' - B_r')\tilde{z}_{ij} + B_r'\sum_{j'}\tilde{z}_{ij'} - \tfrac{1}{Nn}\sum_{hk}(E_r' - B_r')\tilde{z}_{hk} - \tfrac{1}{Nn}\sum_{h,k,k'} B_r'\tilde{z}_{hk'} \\ &= (E_r' - B_r')\zeta_{ij} + B_r'\sum_{j'}\zeta_{ij'}, \end{aligned}$$

so starting from $\alpha = Q^{-1}\tilde{z}$ gives $\tilde{a} = Q^{-1}\zeta$.

References

J. Cai and R. L. Prentice, "Estimating equations for hazard ratio parameters based on correlated failure time data," *Biometrika* vol. 82 pp. 151–164, 1995.

J. Cai and R. L. Prentice, "Regression estimation using multivariate failure time data and a common baseline hazard function model," *Lifetime Data Anal.* vol. 3 pp. 197–213, 1997.

D. Clayton, "A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika* vol. 65 pp. 141–151, 1978.

D. Clayton and J. Cuzick, "Multivariate generalizations of the proportional hazards model (with discussion)," *J. R. Statist. Soc. A* vol. 148 pp. 82–117, 1985.

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

C. Genest and R. J. MacKay, "The joy of copulas: bivariate distributions with uniform marginals," *American Statistician* vol. 40 pp. 280–283, 1986.

E. W. Lee, L. J. Wei, and D. A. Amato, "Cox-type regression analysis for large numbers of small groups of correlated failure time observations," in *Survival Analysis: State of the Art.*, J. P. Klein and P. K. Goel (eds.), Kluwer Academic Publishers, 237–247, 1992.

K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika* vol. 73 pp. 13–22, 1986.

A. W. Marshall and I. Olkin, "Families of Multivariate Distributions,"*J. Amer. Statist. Assoc.* vol. 83 pp. 834–841, 1988.

D. Oakes, "A model for association in bivariate survival analysis," *J. R. Statist. Soc. B* vol. 44 pp. 414–422, 1982.

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

R. L. Prentice and L. Hsu, "Regression on hazard ratios and cross ratios in multivariate failure time analysis," *Biometrika* vol. 84 pp. 349–363, 1997.

C. R. Rao and S. Mitra, *Generalized Inverse of Matrices and its Applications*, Wiley: New York, 1971.

L. J. Wei, D. Y. Lin, and L. Weissfeld, "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *J. Amer. Statist. Assoc.* vol. 84 pp. 1065–1073, 1989.