Statistica Sinica **20** (2010), 1025-1041

ANALYSIS OF MULTIVARIATE FAILURE TIME DATA USING MARGINAL PROPORTIONAL HAZARDS MODEL

Ying Chen, Kani Chen and Zhiliang Ying

Shanghai University of Finance and Economics,
Hong Kong University of Science and Technology and Columbia University

Abstract: The marginal proportional hazards model is an important tool in the analysis of multivariate failure time data in the presence of censoring. We propose a method of estimation via the linear combinations of martingale residuals. The estimation and inference procedures are easy to implement numerically. The estimation is generally more accurate than the existing pseudo-likelihood approach: the size of efficiency gain can be considerable in some cases, and the maximum relative efficiency in theory is infinite. Consistency and asymptotic normality are established. Empirical evidence in support of the theoretical claims is shown in simulation studies.

Key words and phrases: Alternating projection, counting process martingale, marginal likelihood, martingale residual, semiparametric efficiency.

1. Introduction

Multivariate failure time data are common in biomedical studies, engineering, and financial economics. A key feature of this type of data is that the failure times may be related to each other. To analyze the dependence of the failure times on certain covariates, Wei, Lin and Weissfeld (1989) proposed to use a marginal proportional hazards (MPH) model and provided an estimation and inference procedure. The MPH model does not impose any assumption on the interdependence among the multivariate failure times and therefore is quite flexible. Moreover, it inherits many advantages of the well-known proportional hazards model (Cox (1972)). The aim of this paper is to propose a general method of estimation and inference for the MPH model.

Because of the importance of the MPH model and the growing amount of multivariate failure time data, there is an increasing demand for appropriate data analysis. Extensive studies on related subjects are reported in the literature; see Wei, Lin and Weissfeld (1989), Cai and Prentice (1995, 1997), Gray and Li (2002), Oakes (1992, 1997), Lee, Wei and Amato (1992), Prentice and Hsu (1997), Pepe and Cai (1993), Hughes (1995), Therneau (1997), and Yang and Ying (2001), among others. In particular, the estimation method considered in

Wei, Lin and Weissfeld (1989) is based on a pseudo-likelihood that is a product of marginal partial likelihoods. The method is conceptually clear, numerically simple, and easy to implement. However, the pseudo-likelihood approach does not best capture the interdependence of multivariate failure times and may not produce the most accurate estimation of regression parameters. In fact, for the MPH model considered in this paper, the pseudo-likelihood estimation can be significantly improved in some cases. Alternative estimation methods are also proposed and analyzed in Cai and Prentice (1995, 1997) and Gray and Li (2002).

This paper provides a general estimation and inference procedure for the MPH model with common regression parameters across all marginal models. A key issue here is how to properly utilize the interdependence of the failure times to obtain more accurate estimation. For, by the nature of the MPH model, no such dependence structure is available in the model assumptions. To tackle this problem, we propose to use an optimal linear combination of certain martingale residuals that are generated from marginal models. The optimal linear combination coincides with the construction of the quasi-likelihood score (Godambe and Heyde (1987)). The resulting estimating function has a closed form expression and the variance estimator for the regression parameter is easy to obtain as minus the derivative of the estimating function. The estimators are generally more accurate than the maximum pseudo-likelihood estimator. Since the considered estimating functions are in a class of martingale transformations, we can use counting process martingale theory to prove the consistency and asymptotic normality of the estimators. The efficiency improvement over the pseudo-likelihood approach of Wei, Lin and Weissfeld (1989) can be considerable. We raise an example to show that, in theory, the maximum relative efficiency is infinite. Supporting evidence can be found in the simulation studies presented in this paper. Although the proposed method involves partitions that may be too flexible, in view of the possibility of significant efficiency gain it may be worthwhile to explore with even more computational work.

Let $(T_k, C_k, Z_k), 1 \leq k \leq K$, denote the K-variate failure times, censoring times, and the covariates of p dimensions. Set $\mathbf{Z} = (Z_1, \ldots, Z_K)$ and assume Z_k does not concentrate on any $p-1$ dimension hyperplane for some $1 \leq k \leq K$. For every $k = 1, \ldots, K$, T_k and C_k are assumed conditionally independent given $\mathbf{Z}$. The MPH model assumes, for $1 \leq k \leq K$, the hazard function of T_k given $\mathbf{Z}$ satisfies

$$\lambda_{T_k}(t|\mathbf{Z}) = e^{\beta' Z_k} \lambda_k(t), \qquad t \geq 0, \tag{1.1}$$

where β and $\lambda_k(\cdot)$ represent, respectively, the p-dimensional regression parameter and the baseline hazard function. Wei, Lin and Weissfeld (1989) also considered a slightly different model, in which

$$\lambda_{T_k}(t|\mathbf{Z}) = e^{\beta_k' Z_k} \lambda_k(t), \qquad t \geq 0, \qquad k = 1, \ldots, K. \tag{1.2}$$

Here each marginal model has its own regression parameters while model (1.1) has common regression parameters across all K marginal models. A main feature of (1.1) is that the covariate effects on the failures in all marginal models are common and are jointly evaluated.

Model (1.1) can be used in economics, engineering and biomedical studies. For example, it can be applied to the analysis of panel data in econometric studies, e.g., Horowitz and Lee (2004). In finance, it can be used for analysis of time-to-default for closely connected companies. It can also be applied to the evaluation of treatment effects for recurrent diseases in biomedical studies under certain conditions (e.g., specifying, among other conditions, a priori the number of recurrences of interest) or in system reliability in engineering experiments involving multiple components.

We note that there is considerable research devoted to modeling and estimating the dependence structure of multivariate failure times. For example, Bandeen-Roche and Liang (1996) presented a frailty model to capture multi-level dependence of failure times, which is a natural generalization of the ordinal frailty model. Bandeen-Roche and Liang (2002) addressed conditional hazard ratio for multivariate failure times with competing risks; see also Clayton (1978) and Clayton (1985). These studies/methods are different from ours in the absence of modeling of within cluster failure time dependence.

The MPH model with common regression parameters is introduced in Section 2, along with the relevant notation. Section 3 describes the estimation and inference procedures based on a linear combination of martingale residuals. A large sample theory and an argument about the maximum size of efficiency gain are given in Section 4. Simulation studies and an example are presented in Section 5. All proofs are provided in the supplemental material.

2. Pseudo-Partial Likelihood Estimation

With the presence of right censoring, the K event times and their failure/censoring indices are denoted by $Y_k = \min(T_k, C_k)$ and $\delta_k = I(T_k \le C_k)$, $1 \le k \le K$. The observations are n independent and identically distributed (i.i.d.) copies of $(Y_k, \delta_k, Z_k), 1 \le k \le K$, denoted by $(Y_{ik}, \delta_{ik}, Z_{ik}), 1 \le k \le K, 1 \le i \le n$. Throughout, the subscript i or j indicates the i or j-th observation and the subscript k or l indicates the k or l-th event, respectively, and therefore the range of k and l is always $\{1, \ldots, K\}$. Set $Y_k(t) = I(Y_k \ge t)$ and $N_k(t) = \delta_k I(Y_k \le t)$ for $t \ge 0$. Let $M_k(t;\beta) = N_k(t) - \int_0^t e^{\beta Z_k} Y_k(s)\lambda_k(s)ds$. The pseudo-partial likelihood (Wei, Lin and Weissfeld (1989)) is

$$L(\beta) \equiv \prod_{k=1}^{K}\prod_{i=1}^{n}\{\frac{e^{\beta' Z_{ik}}}{\sum_{j=1}^{n} e^{\beta' Z_{jk}} I(Y_{jk} \ge Y_{ik})}\}^{\delta_{ik}}.$$

Here the pseudo-partial likelihood means the product of the Cox partial likelihood for each marginal model, see e.g., Clegg, Cai and Sen (1999). Let $\hat\beta_I$ be the maximizer of $L(\cdot)$. Then, $\hat\beta_I$ is the solution of

$$U(\beta) \equiv \sum_{k=1}^{K} U_k(\beta) = 0, \tag{2.1}$$

where

$$U_k(\beta) = \sum_{i=1}^{n} \int_0^{\tau_k} [Z_{ik} - \bar Z_k(t;\beta)] dN_{ik}(t) = \sum_{i=1}^{n} \int_0^{\tau_k} [Z_{ik} - \bar Z_k(t;\beta)] dM_{ik}(t;\beta), \tag{2.2}$$

$$\tau_k = \sup\{t : P(Y_k > t) > 0\}, \text{ and } \bar Z_k(t;\beta) = \frac{\sum_{i=1}^{n} Z_{ik}\, e^{\beta' Z_{ik}} Y_{ik}(t)}{\sum_{i=1}^{n} e^{\beta' Z_{ik}} Y_{ik}(t)},\ 1 \le k \le K.$$

Here the $U_k(\cdot)$ are the Cox partial likelihood score for the marginal models.

For the model at (1.2), the pseudo-partial likelihood is the same as $L(\cdot)$ except with β replaced by β_k. The estimation and inference for each β_k pertains only to the k-th marginal model, and Cox's partial likelihood procedure for univariate proportional hazards model applies. Simultaneous inference for $\beta_1, \ldots, \beta_K$ requires taking into account the interdependence of multivariate failure and censoring times.

If $(T_k, C_k), k = 1, \ldots, K$, are conditionally independent of each other given the covariates, $\hat\beta_I$ can be shown to be semiparametric efficient. However, this is not the case in general and there exist estimators more accurate than $\hat\beta_I$. In fact, with the simple idea of constructing quasi-likelihood score (Godambe and Heyde (1987)), an estimator more accurate than $\hat\beta_I$ can be found. Specifically, set $\mathbf{U}_o(\beta) = [U_1'(\beta), \ldots, U_K'(\beta)]'$ and let $\mathbf{A}_o = (1/n)(\partial/\partial\beta)\mathbf{U}_o(\beta)|_{\beta=\hat\beta_I}$ and $\mathbf{V}_o = (v_{kl})_{1\le k,l \le K}$, where

$$v_{kl} = \frac{1}{n}\sum_{i=1}^{n} \Big\{ \int_0^{\tau_k} [Z_{ik} - \bar Z_k(t;\beta)] d\tilde M_{ik}(t;\beta) \int_0^{\tau_l} [Z_{il} - \bar Z_l(t;\beta)]' d\tilde M_{il}(t;\beta) \Big\}\Big|_{\beta=\hat\beta_I}, \tag{2.3}$$

with $\tilde M_{ik}$ as defined in Section 3. Notice that $\mathbf{U}_o$ is a pK-vector, $\mathbf{A}_o$ is a $pK \times p$ matrix, v_{kl} are $p \times p$ matrices, and $\mathbf{V}_o$ is a $pK \times pK$ matrix. Consider the estimating equation

$$U_o(\beta) \equiv \mathbf{A}_o' \mathbf{V}_o^{-1} \mathbf{U}_o(\beta) = 0, \tag{2.4}$$

and let $\hat\beta_{II}$ be the solution. Let $D = (I_p, \ldots, I_p)'$ be a $pK \times p$ matrix, where I_p is the $p \times p$ identity matrix. Then, under regularity conditions, the asymptotic variances of $\hat\beta_I$ and $\hat\beta_{II}$ are $(D'\mathbf{A}_o)^{-1}(D'\mathbf{V}_o D)(\mathbf{A}_o' D)^{-1}$ and $(\mathbf{A}_o'\mathbf{V}_o^{-1}\mathbf{A}_o)^{-1}$, respectively, and the latter is smaller than or equal to the former in the sense of nonnegative definiteness. The consistency of the variance estimators is shown in Wei, Lin and Weissfeld (1989), and they point out the important fact that $\mathbf{V}_o$ is a consistent estimator of the covariance matrix of $\mathbf{U}_o$. A general asymptotic theory is presented in Section 4.

The construction of $U_o(\cdot)$ is a result of the optimal linear combination of estimating functions. This simple procedure may be traced back to the classical theory of quasi-likelihood (Godambe and Heyde (1987)), and it has been used extensively in the literature: see Liang and Zeger (1986), Qu and Lindsay (2003), and Fine, Yan and Kosorok (2004), among others. In general, for a class of estimating functions $\psi_1(\beta), \ldots, \psi_m(\beta)$ that are conditionally mean zero, the optimal linear combination of the estimating functions is $A'V^{-1}\Psi$, where $\Psi = (\psi_1', \ldots, \psi_m')'$, V is the conditional variance matrix of Ψ, and A is the conditional mean of $(\partial/\partial\beta)\Psi$. Under certain regularity conditions on the large sample behavior of $\psi_1, \ldots, \psi_m$, the estimator based on this estimating function has the smallest asymptotic variance among all estimators based on linear combinations of Ψ. This procedure is simple, yet effective. It is an important ingredient in the methodology of generalized estimating equations (Liang and Zeger (1986)). More sophisticated estimators derived in the next section are based on this procedure.

3. Estimation Method Based on Linear Combinations of Martingale Residuals

In this section, we consider a class of estimating functions that can be viewed as linear combinations of martingale residuals and apply the optimal linear combination procedure to obtain estimators with improved accuracy. Recall that U and U_o are linear combinations of the U_k which are Cox's partial likelihood scores for univariate proportional hazards models. In the case of MPH model, however, using U_k as the building blocks to construct estimating functions may be rather restrictive, since the U_k may contain insufficient information about the interdependence of the multivariate failure and censoring times.

To overcome this difficulty, we consider linear combinations of the martingale differences $dM_{ik}(t;\beta)$. Since the $dM_{ik}(t;\beta)$ contain the unknown baseline functions, they cannot be directly used as building blocks. Instead we consider linear combination of the martingale residuals

$$\begin{aligned} d\tilde{M}_{ik}(t;\beta) &\equiv dN_{ik}(t) - e^{\beta' Z_{ik}} Y_{ik}(t) \frac{\sum_{j=1}^n dN_{jk}(t)}{\sum_{j=1} e^{\beta' Z_{jk}} Y_{jk}(t)} \\ &= dM_{ik}(t;\beta) - e^{\beta' Z_{ik}} Y_{ik}(t) \frac{\sum_{j=1}^n dM_{jk}(t;\beta)}{\sum_{j=1} e^{\beta' Z_{jk}} Y_{jk}(t)} \end{aligned}$$

that are free of the baseline functions. In this broad class of estimating functions, more accurate estimators can be found. Notice that, in the case of univariate proportional hazards model, it can be shown that the optimal linear combination of these martingale residuals is identical to the Cox partial likelihood score, as a special case of the Hutton-Nelson solution.

Let $\mathbf{h} \equiv (h_1, \ldots, h_K)$, where each h_k is a p-dimensional measurable function of $(t, \mathbf{Z})$ defined on $[0, \infty) \times R^{pK}$, and set $h_{ik}(t) = h_k(t, \mathbf{Z}_i)$. Consider the estimating function

$$\begin{aligned} U_{\mathbf{h}}(\beta) &\equiv \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^{\tau_k} h_{ik}(t) d\tilde{M}_{ik}(t;\beta) \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^{\tau_k} [h_{ik}(t) - \bar{h}_k(t;\beta)] dN_{ik}(t) \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^{\tau_k} [h_{ik}(t) - \bar{h}_k(t;\beta)] dM_{ik}(t;\beta), \end{aligned} \tag{3.1}$$

where

$$\bar{h}_k(t;\beta) = \frac{\sum_{i=1}^n h_{ik}(t) e^{\beta' Z_{ik}} Y_{ik}(t)}{\sum_{i=1}^n e^{\beta' Z_{ik}} Y_{ik}(t)}.$$

Let $\hat{\beta}_{\mathbf{h}}$ be the solution of $U_{\mathbf{h}}(\beta) = 0$. It is seen from (3.1) that $U_{\mathbf{h}}$ can be viewed as a linear combination of the martingale residuals $d\tilde{M}_{ik}$.

The choice of $\mathbf{h}$ determines the accuracy of the resulting estimator. Choosing $h_{ik}(t) = Z_{ik}$, $U_{\mathbf{h}}(\cdot)$ reduces to $U(\cdot)$ in (2.1) and gives rise to the estimator $\hat{\beta}_I$. Choosing $h_{ik}(t) = W_k Z_{ik}$ with the optimal $p \times p$ matrix W_k, $U_{\mathbf{h}}(\cdot)$ reduces to $U_o(\cdot)$ in (2.3) and gives rise to the estimator $\hat{\beta}_{II}$. More accurate estimation of β requires more computational load in the choice of $\mathbf{h}$. Using the idea of optimal linear combination presented in Section 2, we propose to use the following estimating functions as building blocks to construct optimal linear combination.

Let $\mathcal{A}_1, \ldots, \mathcal{A}_m$ be a partition of the space $[0, \infty) \times R^{pK}$. For $1 \le s \le m$ and $1 \le k, l \le K$, let

$$g_{iks}(t) = I\{(t, \mathbf{Z}_i) \in \mathcal{A}_s\} Z_{ik}, \qquad \bar{g}_{ks}(t;\beta) = \frac{\sum_{j=1}^n g_{jks}(t) e^{\beta' Z_{jk}} Y_{jk}(t)}{\sum_{j=1}^n e^{\beta' Z_{jk}} Y_{jk}(t)},$$

$$u_{ks}(\beta) = \sum_{i=1}^{n} \int_0^{\tau_k} g_{iks}(t) d\tilde{M}_{ik}(t;\beta) = \sum_{i=1}^{n} \int_0^{\tau_k} [g_{iks}(t) - \bar{g}_{ks}(t;\beta)] dN_{ik}(t;\beta).$$

Similar to the construction of U_o, define $\mathbf{U}(\beta) = (u'_{11}, \ldots, u'_{1m}, \ldots, u'_{K1}, \ldots, u'_{Km})'$, a pKm-vector. Similar to (2.3), the covariance of u_{ks} and u_{ls*} can be estimated by

$$\sum_{i=1}^{n} \Big\{ \int_0^{\tau_k} [g_{iks}(t) - \bar{g}_{ks}(t;\beta)] d\tilde{M}_{ik}(t;\beta) \int_0^{\tau_l} [g_{ils^*}(t) - \bar{g}_{ls^*}(t;\beta)]' d\tilde{M}_{il}(t;\beta) \Big\} \Big|_{\beta = \hat{\beta}_{II}}.$$

Let $\mathbf{V}$ be the estimated variance matrix of $\mathbf{U}$. Let $\mathbf{A} = (\partial/\partial\beta)\mathbf{U}(\beta)|_{\beta=\hat{\beta}_{II}}$ and consider the estimating equation $\mathbf{A}'\mathbf{V}^{-1}\mathbf{U}(\beta) = 0$, with $\hat{\beta}$ be the solution. The standard quasi-likelihood procedure implies that this estimating function is optimal among all linear combinations of $\mathbf{U}$. Moreover, the asymptotic variance of $\hat{\beta}$ is consistently estimated by $(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}$.

This estimation method offers estimators with smaller asymptotic variance. If the partition is the trivial one, i.e., $m = 1$ and $\mathcal{A}_1$ is the entire space, $\hat{\beta}$ is identical to $\hat{\beta}_{II}$. In general, the finer the partition, the more accurate the estimation, since there are more "building blocks" for constructing the optimal linear combination. In theory, $\hat{\beta}$ can approximate the semiparametric efficient estimator when the partition is fine enough, but the partition cannot be too fine given a finite sample. Still, the computation of the proposed estimator $\hat{\beta}$ and its variance estimator is straightforward. In view of the improved estimation accuracy even with a crude partition, the proposed procedure deserves study.

4. An Asymptotic Theory

The asymptotic properties of our estimator can be established following the counting process martingale theory developed by Andersen and Gill (1982) and others. Some notation is needed. Let β_0 and $\lambda_{k0}(\cdot)$ be the true β and $\lambda_k(\cdot)$, respectively. Let $\mu_{k,\mathbf{h}}(t) = E(h_k(t,\mathbf{Z})|Y_k = t, \delta_k = 1)$, $\mu_k(t) = E(Z_k|Y_k = t, \delta_k = 1)$, $M_k(t) = N_k(t) - \int_0^t e^{\beta_0' Z_k} Y_k(s)\lambda_{k0}(s)ds$, $\xi_{k,\mathbf{h}} = \int_0^{\tau_k}[h_k(t,\mathbf{Z}) - \mu_{k,\mathbf{h}}(t)]dM_k(t)$ and $\xi_k = \int_0^{\tau_k}[Z_k - \mu_k(t)]dM_k(t)$.

Similar to Theorem 8.4.1 of Fleming and Harrington (1991, p.305) and Andersen and Gill (1982), we set the following regularity conditions. Throughout, $\otimes 2$ denotes outer product.

(C1) The components of $\mathbf{h}$ are bounded and measurable functions of $(t, \mathbf{Z})$;

(C2) $\mathbf{Z}$ is bounded;

(C3) $\tau_k < \infty$, $\int_0^{\tau_k}\lambda_{k0}(t)dt < \infty$, and $\lambda_{k0}(t)$ is continuous on $[0, \tau_k]$, for $k = 1, \ldots, K$;

(C4) $A_{\mathbf{h}} \equiv \sum_{k=1}^K E(\xi_{k,\mathbf{h}}\xi_k')$ and $V_{\mathbf{h}} \equiv E\Big[\Big(\sum_{k=1}^K \xi_{k,\mathbf{h}}\Big)^{\otimes 2}\Big]$ are both finite and non-degenerate.

Theorem. *Under* (C1)$-$(C4), *there exists a solution,* $\hat{\beta}_{\mathbf{h}}$, *of* $U_{\mathbf{h}}(\beta) = 0$ *such that* $n^{1/2}(\hat{\beta}_{\mathbf{h}} - \beta_0) \to N(0, \Sigma_{\mathbf{h}})$, *where* $\Sigma_{\mathbf{h}} = A_{\mathbf{h}}^{-1}V_{\mathbf{h}}{A_{\mathbf{h}}'}^{-1}$.

If $K = 1$, or if (Y_k, δ_k), $1 \le k \le K$, are conditionally independent given $\mathbf{Z}$, then $\Sigma_{\mathbf{h}} = [\sum_{k=1}^K E(\xi_k^{\otimes 2})]^{-1}$ when $h_k(\mathbf{Z}, t) = Z_k$ for $k = 1, \ldots, K$. We note that the estimators $\hat{\beta}_I$, $\hat{\beta}_{II}$, and $\hat{\beta}$ are all special cases of $\hat{\beta}_{\mathbf{h}}$ with different choices of $\mathbf{h}$.

Proposition. *Under* (C2)−(C3), *there exists an* $\mathbf{h}^\star$*, satisfying* (S2.1) *in the Supplement, such that* $\sum_{k=1}^{K}\xi_{k,\mathbf{h}^\star}$ *is the efficient score and* $\Sigma_{\mathbf{h}^\star}^{-1}$ *is the minimum Fisher information for* β*.*

The proposition establishes $\mathbf{h}^\star$ so that $\Sigma_{\mathbf{h}^\star}$ achieves a minimum over $\mathbf{h}$. Analytically, the minimization can be expressed in terms of alternating projections, as shown in (S2.1) in the Supplement. Notice that $U_{\mathbf{h}}$, $\hat{\beta}_{\mathbf{h}}$ and $\Sigma_{\mathbf{h}}$ are unchanged if $\mathbf{h}$ is changed by adding any function of t. The unique solution $\mathbf{h}^\star$ can be expressed as the solution to

$$h_k(\mathbf{Z},t)+\sum_{\substack{1\le l\le K\\ l\neq k}}\int_0^{\tau_l}[h_l(\mathbf{Z},s)-\mu_{l,\mathbf{h}}(s)]\frac{\Gamma_{l,k}(s,t|\mathbf{Z})}{\Gamma_k(t|\mathbf{Z})}ds=Z_k,\ t\in[0,\tau_k),\ 1\le k\le K,$$

where $\Gamma_{l,k}(s,t|\mathbf{Z})=E(dM_l(s)dM_k(t)|\mathbf{Z})/(dsdt)$ and $\Gamma_l(t|\mathbf{Z})=E(dM_l(t)dM_l(t)|\mathbf{Z})/dt=P(Y_l\ge t|\mathbf{Z})e^{\beta_0' Z_l}\lambda_{l0}(t)$, for $1\le l\neq k\le K$. These equations can be derived from (S2.1) or through direct minimization of $\Sigma_{\mathbf{h}}$. A natural way to obtain semiparametric efficient estimation appears to be by solving for $\mathbf{h}^\star$. Unfortunately, this involves the estimation of $\Gamma_{l,k}(s,t|\mathbf{Z})$ and $\Gamma_l(t|\mathbf{Z})$, and the curse of dimensionality intrudes. Moreover, it is difficult to assess the stability of the numerical solutions, especially when the Γ functions are replaced by estimators that are possibly inaccurate. Our experience with numerical studies indicates that this approach, although conceptually clear, is numerically difficult to carry out.

The estimation approach that we propose is easy to implement computationally. In view of the difficulty of the direct use of $\mathbf{h}^\star$, it is worthwhile to consider optimization over step functions over a partition $\mathcal{A}_1,\ldots,\mathcal{A}_m$, i.e., each component of $\mathbf{h}$ is constant over each cell of the partition. The optimization is carried out by the "variance-inverse-derivative-transpose" procedure as discussed in Sections 2 and 3. The resulting optimal estimator has asymptotic variance achieving the minimum of $\Sigma_{\mathbf{h}}$ over all $\mathbf{h}$ that are step functions over the partition. In contrast, the variance of semiparametric efficient estimation achieves the minimum of $\Sigma_{\mathbf{h}}$ over all $\mathbf{h}$. Since measurable functions can be closely approximated by step functions, $\Sigma_{\mathbf{h}^\star}$ can be closely approximated by $\Sigma_{\mathbf{h}}$ with $\mathbf{h}$ the optimal step function over a partition that is fine enough, and $\mathbf{h}^\star$ is closely approximated by this $\mathbf{h}$.

The proposition establishes the semiparametric efficiency of the estimator $\hat{\beta}_{\mathbf{h}^\star}$, and an example shows that there can be large efficiency gain over the estimators $\hat{\beta}_I$ or $\hat{\beta}_{II}$. For simplicity, consider $p=1$, $K=2$, and any estimator, denoted by $\hat{\beta}_w$, as the solution of

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau_k}[Z_{ik}-\bar{Z}_k(t;\beta)]w_k(t)dN_{ik}(t)=0,$$

where $w_k(\cdot)$ is a deterministic weight function. In theory, the maximum size of efficiency improvement of $\hat{\beta}_{\mathbf{h}^\star}$ over the optimal $\hat{\beta}_w$ can be infinitely large, and $\hat{\beta}_I$ and $\hat{\beta}_{II}$ are two special cases of $\hat{\beta}_w$. Consider that $\beta_0 = 0$, $C_1 = C_2 = \infty$. Let Z_1, Z_2, T_1 and ϵ be independent, all uniform on $[0, 1]$. Take $T_2 = T_1 + a\epsilon$ where $a > 0$ is a constant. Choose $\mathbf{h} = (h_1, h_2)$ with $h_1 = -h_2 = (Z_1 - Z_2)$. Observe that $T_2 \to T_1$ as $a \to 0$. This implies that $E(\xi_{1,\mathbf{h}} + \xi_{2,\mathbf{h}})^2 \to 0$ and $E(\xi_{1,\mathbf{h}}\xi_1 + \xi_{2,\mathbf{h}}\xi_2) \to E(Z_1 - Z_2)^2 = 1/6$ as $a \to 0$. Recall that $\Sigma_{\mathbf{h}} = [E(\xi_{1,\mathbf{h}}\xi_1 + \xi_{2,\mathbf{h}}\xi_2)]^{-2}E(\xi_{1,\mathbf{h}} + \xi_{2,\mathbf{h}})^2$. Therefore, as $a \to 0$, $\Sigma_{\mathbf{h}} \to 0$, and it follows that $\Sigma_{\mathbf{h}^\star} \to 0$ since $\Sigma_{\mathbf{h}^\star} \leq \Sigma_{\mathbf{h}}$. On the other hand, as $a \to 0$, the asymptotic variance of $\hat{\beta}_w$ converges to

$$\frac{(1/12)\int_0^1 [w_1^2(t)) + w_2^2(t)]dt}{\{(1/12)\int_0^1 [w_1(t) + w_2(t)]dt\}^2} \geq 6,$$

where the equality holds if and only if $w_1(t) = w_2(t) = c$ for some nonzero constant c. Then the relative efficiency of the estimator $\hat{\beta}_{\mathbf{h}^\star}$ with respect to $\hat{\beta}_w$ can be arbitrarily large when a is arbitrarily close to 0. Heuristically, when the T_k are strongly dependent on each other and the Z_k are not, the size of efficiency improvement can be large. In our extreme case, $T_2 = T_1$, β can be identified as 0 given finite sample size.

5. Simulation Studies and Example

Extensive Monte Carlo studies with sample size $n = 200$ were carried out to examine the finite sample properties of the proposed estimation method. In all four simulation examples, K= 2, and the two baseline hazard functions were constant at 1. The maximum correlation coefficient between the failure times T_1 and T_2, from the Gumbel (1960) distribution used in Wei, Lin and Weissfeld (1989), was fixed as 0.25. In our simulation, we used the Weighted Linear Combination (WLC) method (Johnson and Tenenbein (1981)) to generate the failure times T_1 and T_2 that marginally are univariate exponentials with hazard rates $\exp(\beta Z_1)$ and $\exp(\beta Z_2)$, respectively. Using the WLC method, the degree of dependence between T_1 and T_2 is determined by the parameter ρ_s, Spearman's rho with the range [0, 1], and distinct from the correlation between T_1 and T_2.

In Example I, the covariates Z_1 and Z_2 were independent of each other, with $P(Z_i = 0) = P(Z_i = 1)$= 1/2, $i = 1, 2$. Spearman's rho was set at 0.50; the regression parameter β took values $-0.5, -0.4, ..., 0.4, 0.5$; the censoring variables $C_1 = C_2$ were set to be constant at 2. The proportion of censoring is generally moderate, for example, it is about 29.73% when β = 0.5 and Z_1= 1. To compute the proposed estimator $\hat{\beta}$, we chose to partition R^{pK} into four cells: $\mathcal{A}_1 = \{Z_1 \leq 0.5, Z_2 \leq 0.5\}$, $\mathcal{A}_2 = \{Z_1 \leq 0.5, Z_2 > 0.5\}$,

Table 1. Simulation results for Example I

β	$\hat{\beta}_I$	$\hat{\beta}_{II}$	$\hat{\beta}_{III}$	$\hat{\beta}$	$Var(\hat{\beta}_I)$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta}_{II})}$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta}_{III})}$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta})}$	ECP by $\hat{\beta}$ (99%)
-0.50	-0.498	-0.498	-0.493	-0.505	0.0129	1.000	1.005	1.155	0.989
-0.40	-0.392	-0.392	-0.400	-0.398	0.0132	1.000	1.004	1.182	0.986
-0.30	-0.297	-0.297	-0.291	-0.300	0.0125	1.000	1.005	1.150	0.984
-0.20	-0.199	-0.199	-0.187	-0.203	0.0112	1.000	1.006	1.149	0.986
-0.10	-0.105	-0.105	-0.098	-0.107	0.0114	1.000	1.006	1.137	0.981
0	0.002	0.002	0.009	0.000	0.0125	1.000	1.006	1.225	0.982
0.10	0.106	0.106	0.107	0.109	0.0114	1.000	1.006	1.164	0.985
0.20	0.199	0.199	0.202	0.202	0.0114	1.000	1.007	1.198	0.989
0.30	0.304	0.304	0.297	0.305	0.0116	1.000	1.008	1.192	0.984
0.40	0.400	0.400	0.402	0.406	0.0118	1.000	1.008	1.127	0.987
0.50	0.502	0.502	0.505	0.509	0.0110	1.000	1.010	1.121	0.979

Notes: 1. $\hat{\beta}_I$: the WLW estimator proposed by Wei, Lin and Weissfeld (1989).
2. $\hat{\beta}_{II}$: the optimal WLW estimator at (2.4).
3. $\hat{\beta}_{III}$: the weighted estimator proposed by Cai and Prentice (1995).
4. $\hat{\beta}$: the proposed estimator.
5. ECP: empirical coverage probability of the $(1-\alpha)\%$ confidence interval for β.

$\mathcal{A}_3 = \{Z_1 > 0.5, Z_2 \leq 0.5\}$, and $\mathcal{A}_4 = \{Z_1 > 0.5, Z_2 > 0.5\}$. The time horizon was not partitioned. To compare with $\hat{\beta}$, the estimators $\hat{\beta}_I$, $\hat{\beta}_{II}$, and $\hat{\beta}_{III}$ were also calculated using Newton-Raphson iterative procedures. Here $\hat{\beta}_{III}$ is the weighted estimator proposed by Cai and Prentice (1995) where the weight matrix is estimated using the nonparametric method proposed by Prentice and Cai (1992). The simulation results are presented in Table 1, which shows the average of the estimates, empirical variances of the WLW estimator $\hat{\beta}_I$, empirical relative efficiencies for $\hat{\beta}_{II}$, $\hat{\beta}_{III}$, and $\hat{\beta}$ over $\hat{\beta}_I$, and empirical coverage probabilities(ECP) of the 99% confidence intervals for β using the proposed method. All simulation results were based on 1,000 replications except for $\hat{\beta}_{III}$, which was based on 500 simulation runs. As shown in the table, no efficiency improvement of $\hat{\beta}_{II}$ over $\hat{\beta}_I$ existed, the relative efficiencies for $\hat{\beta}_{III}$ over $\hat{\beta}_I$ were negligible, the efficiency improvement of the proposed method over $\hat{\beta}_I$ was moderate, and the empirical coverage probabilities of the intervals using the proposed method were close to the nominal confidence levels. The reason for the moderate efficiency gain of the proposed method was likely due to the crude partition and the weak correlation between T_1 and T_2.

Example II was designed to verify the claim that the maximum efficiency improvement is, in theory, infinity. The setup was the same as that of Example I except that we took ρ_s from 0.9 to 0.9999. The partition for calculating $\hat{\beta}$ was also the same as Example I. Table 2 reports the results for $\hat{\beta}_{III}$ and $\hat{\beta}$ with ρ_s=0.9, 0.99, 0.999, 0.9999, and β=0. The simulation results in the absence of censorship are also listed in the table, where the size of efficiency gain of $\hat{\beta}_{III}$ over $\hat{\beta}_I$ can be

Table 2. Simulation results for Example II.

Censoring	ρ_s	$Corr(T_1, T_2)$	$\hat{\beta}_{III}$	$\hat{\beta}$	$Var(\hat{\beta}_I)$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta}_{III})}$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta})}$
no	0.9000	0.781	0.061	0.002	0.0104	1.339	2.336
	0.9900	0.932	0.072	0.000	0.0102	1.935	6.818
	0.9990	0.979	0.078	0.000	0.0101	2.618	28.756
	0.9999	0.994	0.078	-0.000	0.0111	3.211	134.342
c=2	0.9000	0.781	0.011	0.004	0.0117	1.013	2.608
	0.9900	0.932	0.009	0.002	0.0119	1.019	11.162
	0.9990	0.979	0.013	0.000	0.0120	1.021	68.135
	0.9999	0.994	0.008	-0.000	0.0119	1.020	468.231

Notes: 1. $\hat{\beta}_I$: the WLW estimator proposed by Wei, Lin and Weissfeld (1989).
2. $\hat{\beta}_{III}$: the weighted estimator proposed by Cai and Prentice (1995).
3. $\hat{\beta}$: the proposed estimator.

large if T_1 and T_2 are strongly dependent on each other. The relative efficiency of $\hat{\beta}$ over $\hat{\beta}_I$ is much larger than that of $\hat{\beta}_{III}$ over $\hat{\beta}_I$, and it is interesting that in the presence of censorship the efficiency gain of $\hat{\beta}_{III}$ over $\hat{\beta}_I$ decreases quickly while the efficiency gain of $\hat{\beta}$ over $\hat{\beta}_I$ increases. The reason may be that the weights used in $\hat{\beta}_{III}$ are less precisely estimated by the nonparametric estimation. In this example, the variance estimation of the proposed method seemed to be biased upward, but bias decreased as the sample size increased.

In Example III, we considered dependent covariates Z_1 and Z_2, in which Z_1 was uniform on $(0,1)$ and $Z_2 = Z \cdot I(Z > 0)$, where Z was normal with mean $Z_1/4$ and variance 0.05^2, given Z_1. Censoring times C_1 and C_2 were $U(1,3)$ variates, independent of each other and of T_1 and T_2. The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \leq 0.3\}$ and $\mathcal{A}_2 = \{Z_1 > 0.3\}$. The time horizon was not partitioned. Table 3 shows selected results at configurations determined by β-values of $-0.5, 0$, and 0.5, and values of ρ_s= 0.70, 0.80, 0.90, and 0.99. In this example, all simulation results were based on 500 replications. The relative efficiencies here were all larger than those in Example I. Compared with the results in Example I, the efficiency gains for $\hat{\beta}_{III}$ over $\hat{\beta}_I$ ranged over $[1.093, 1.450]$, for example. The largest efficiency improvement over $\hat{\beta}_I$ was provided by $\hat{\beta}$, ranging up to 2.668. Compared with Example I, Example III had dependent covariate structure and fewer partition cells. The dependence of covariates Z_1 and Z_2 contributes more efficiency improvement than that of the partition in this case. Some nonconvergence occurred in using the Newton-Raphson procedure to compute $\hat{\beta}_{III}$, presumably owing to unstable estimated weights under some sampling configurations, while no other nonconvergence occurred to compute the other estimators.

In Example IV, we considered different partitions to see how the relative efficiency of $\hat{\beta}$ over $\hat{\beta}_I$ changes. In this example, the setup was the same as that of

Table 3. Simulation results for Example III.

β	ρ_s	$\hat\beta_{II}$	$\hat\beta_{III}$(nonconvergence)	$\hat\beta$	$Var(\hat\beta_I)$	$\frac{Var(\hat\beta_I)}{Var(\hat\beta_{II})}$	$\frac{Var(\hat\beta_I)}{Var(\hat\beta_{III})}$	$\frac{Var(\hat\beta_I)}{Var(\hat\beta)}$	ECP by $\hat\beta$(99%)
-0.50	0.70	-0.496	-0.506 (15.4%)	-0.505	0.0996	1.243	1.173	1.287	0.982
	0.99	-0.505	-0.524 (16.0%)	-0.514	0.1099	1.981	1.243	2.668	0.992
0	0.70	0.010	-0.007 (23.0%)	0.001	0.0890	1.234	1.211	1.273	0.986
	0.80	0.002	-0.009 (20.0%)	-0.004	0.0819	1.263	1.093	1.342	0.984
	0.90	-0.020	-0.037 (20.2%)	-0.024	0.0895	1.462	1.231	1.601	0.978
	0.99	0.005	-0.025 (18.4%)	0.000	0.0995	1.790	1.176	2.429	0.972
0.50	0.70	0.509	0.488 (33.6%)	0.501	0.0830	1.221	1.192	1.256	0.984
	0.99	0.500	0.454 (30.2%)	0.491	0.0923	1.965	1.450	2.542	0.984

Notes: 1. $\hat\beta_I$: the WLW estimator proposed by Wei, Lin and Weissfeld (1989).

2. $\hat\beta_{II}$: the optimal WLW estimator at (2.4).

3. $\hat\beta_{III}$: the weighted estimator proposed by Cai and Prentice (1995).

4. $\hat\beta$: the proposed estimator.

5. ECP: empirical coverage probability of the $(1-\alpha)$% confidence interval for β.

Example III except for partitions. Five partitions were considered. Partition I: The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \le 0.5\}$ and $\mathcal{A}_2 = \{Z_1 > 0.5\}$, and the time horizon was not partitioned. Partition II: The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \le 0.5\}$ and $\mathcal{A}_2 = \{Z_1 > 0.5\}$, and the time horizon $[0, +\infty)$was partitioned into two intervals so that each had about the same number of observations. Partition III: The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \le 0.25\}$, $\mathcal{A}_2 = \{0.25 < Z_1 \le 0.5\}$, and $\mathcal{A}_3 = \{Z_1 > 0.5\}$, and the time horizon was not partitioned. Partition IV: The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \le 0.25\}$, $\mathcal{A}_2 = \{0.25 < Z_1 \le 0.5\}$, and $\mathcal{A}_3 = \{Z_1 > 0.5\}$, and the time horizon was partitioned into two intervals so that each had about the same number of observations. Partition V: The space R^{pK} was partitioned into $\mathcal{A}_1 = \{Z_1 \le 0.5\}$ and $\mathcal{A}_2 = \{Z_1 > 0.5\}$, and the time horizon $[0, +\infty)$was partitioned into three intervals, each of which contained about the same number of observations.

Table 4 lists the simulation results based on 500 replications. It indicates that the efficiency improvement of $\hat\beta$ is sizable if the degree of dependence between T_1 and T_2, the degree of dependence between Z_1 and Z_2, or/and the number of partition is sufficiently large. Balanced partitions for $\hat\beta$ yielded larger efficiency gains than unbalanced partitions. The efficiency improvement of $\hat\beta$, based on partition, was always larger than that of $\hat\beta_{II}$ and $\hat\beta_{III}$. $\hat\beta_{II}$ possessed larger efficiency gains than $\hat\beta_{III}$ only if the degree of dependence between Z_1 and Z_2 was strong. Additionally, there were some cases in which the efficiency gain of $\hat\beta$ over the WLW method was huge. However, it should be noted that the proposed method needs enough sample to keep the accuracy of the estimated variances, and to obtain large efficiency gains with more partitions.

Finally, we illustrate the proposed method using a data set. The data are the recurrence times between the insertion of a catheter and the next infection

Table 4. Simulation results for Example IV.

Partition	β	ρ_s	$\hat{\beta}$	$Var(\hat{\beta}_I)$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta})}$	ECP by $\hat{\beta}$(99%)
I	-0.50	0.70	-0.505	0.0996	1.297	0.986
		0.80	-0.516	0.0910	1.391	0.986
		0.90	-0.524	0.1008	1.703	0.990
		0.99	-0.512	0.1128	2.415	0.986
	0	0.70	0.000	0.0890	1.299	0.986
		0.80	-0.005	0.0819	1.356	0.984
		0.90	-0.023	0.0895	1.745	0.990
		0.99	-0.005	0.0995	2.673	0.970
	0.50	0.70	0.499	0.0830	1.272	0.980
		0.80	0.498	0.0789	1.386	0.986
		0.90	0.481	0.0816	1.653	0.986
		0.99	0.498	0.0968	2.577	0.970
II	-0.50	0.70	-0.511	0.0996	1.267	0.980
		0.80	-0.515	0.0910	1.419	0.990
		0.90	-0.522	0.1008	1.900	0.992
		0.99	-0.514	0.1128	3.167	0.980
	0	0.70	0.001	0.0890	1.271	0.988
		0.80	0.001	0.0819	1.407	0.990
		0.90	-0.021	0.0895	1.906	0.994
		0.99	-0.003	0.0995	3.192	0.970
	0.50	0.70	0.506	0.0830	1.244	0.980
		0.80	0.508	0.0789	1.397	0.984
		0.90	0.482	0.0816	1.741	0.982
		0.99	0.500	0.0968	2.883	0.970
III	-0.50	0.70	-0.512	0.0996	1.310	0.982
		0.80	-0.521	0.0910	1.365	0.982
		0.90	-0.530	0.1008	1.674	0.982
		0.99	-0.513	0.1128	2.423	0.982
	0	0.70	-0.007	0.0890	1.304	0.984
		0.80	-0.011	0.0819	1.335	0.980
		0.90	-0.031	0.0895	1.667	0.980
		0.99	-0.009	0.0995	2.666	0.968
	0.50	0.70	0.491	0.0830	1.276	0.980
		0.80	0.492	0.0789	1.362	0.988
		0.90	0.471	0.0816	1.573	0.972
		0.99	0.494	0.0968	2.543	0.968
IV	-0.50	0.70	-0.515	0.0996	1.277	0.980
		0.80	-0.520	0.0910	1.373	0.990
		0.90	-0.528	0.1008	1.850	0.988
		0.99	-0.514	0.1128	3.101	0.978
	0	0.70	-0.006	0.0890	1.260	0.984
		0.80	-0.005	0.0819	1.340	0.984
		0.90	-0.028	0.0895	1.799	0.978
		0.99	-0.004	0.0995	3.127	0.966
	0.50	0.70	0.498	0.0830	1.231	0.978
		0.80	0.499	0.0789	1.351	0.982
		0.90	0.473	0.0816	1.614	0.968
		0.99	0.496	0.0968	2.826	0.960

Table 4. (Continuous)

Partition	β	ρ_s	$\hat{\beta}$	$Var(\hat{\beta}_I)$	$\frac{Var(\hat{\beta}_I)}{Var(\hat{\beta})}$	ECP by $\hat{\beta}$(99%)
V	-0.50	0.70	-0.519	0.0996	1.220	0.974
sample size		0.80	-0.521	0.0910	1.388	0.986
$n = 200$		0.90	-0.523	0.1008	1.908	0.988
		0.99	-0.512	0.1128	3.431	0.984
V	-0.50	0.70	-0.519	0.0352	1.287	0.984
sample size		0.80	-0.510	0.0346	1.518	0.994
$n = 500$		0.90	-0.496	0.0414	2.020	0.978
		0.99	-0.501	0.0445	4.117	0.990

Notes: 1. $\hat{\beta}_I$: the WLW estimator proposed by Wei, Lin and Weissfeld (1989).
2. $\hat{\beta}$: the proposed estimator.
3. ECP: empirical coverage probability of the $(1-\alpha)\%$ confidence interval for β.

of kidney patients who were using a portable dialysis machine (McGilchrist and Aisbett (1991)). 38 patients were observed, 10 male. For each patient, two recurrence times Y_1 and Y_2 were recorded. The catheter could be removed for other reasons so that there are some censored data. The data are given in Table 5.

One question is whether there is a difference in recurrence times between male and female. Here the response is recurrence time Y_i, with covariate Z_i as 1 if male, 2 if female. The covariate space R^{pK} was first partitioned into $\mathcal{A}_1 = \{Z_1 \leq 1.5, Z_2 \leq 1.5\}$, $\mathcal{A}_2 = \{Z_1 \leq 1.5, Z_2 > 1.5\}$, $\mathcal{A}_3 = \{Z_1 > 1.5, Z_2 \leq 1.5\}$, and $\mathcal{A}_4 = \{Z_1 > 1.5, Z_2 > 1.5\}$, and the time horizon was not partitioned. The estimate $\hat{\beta}$ of sex effect using the proposed method was -1.327 and the approximate 95% and 99% confidence intervals were $[-2.165, -0.488]$ and $[-2.429, -0.224]$, respectively. Thus the sex effect was significant at the 1% level. To check sensitivity to partitioning, some other partitions were applied. If the space was partitioned into $\mathcal{A}_1 = [0, t_{11}) \times \{Z_1 \leq 1.5, Z_2 \leq 1.5\}$, $\mathcal{A}_2 = [t_{11}, +\infty) \times \{Z_1 \leq 1.5, Z_2 \leq 1.5\}$, $\mathcal{A}_3 = [0, t_{12}) \times \{Z_1 \leq 1.5, Z_2 > 1.5\}$, $\mathcal{A}_4 = [t_{12}, +\infty) \times \{Z_1 \leq 1.5, Z_2 > 1.5\}$, $\mathcal{A}_5 = [0, t_{21}) \times \{Z_1 > 1.5, Z_2 \leq 1.5\}$, $\mathcal{A}_6 = [t_{21}, +\infty) \times \{Z_1 > 1.5, Z_2 \leq 1.5\}$, $\mathcal{A}_7 = [0, t_{22}) \times \{Z_1 > 1.5, Z_2 > 1.5\}$, and $\mathcal{A}_8 = [t_{22}, +\infty) \times \{Z_1 > 1.5, Z_2 > 1.5\}$, and time partitioned so that each subinterval had about the same number of observations, $\hat{\beta} = -1.671$, and the estimated variances of $\hat{\beta}$ was 0.1050. The corresponding approximate 95% and 99% confidence intervals were $[-2.306, -1.036]$ and $[-2.506, -0.837]$, respectively. Again, there was a sex effect. When partitioning time into three intervals with almost the same number of observations, we found $\hat{\beta} = -1.641$, and the estimated variance of $\hat{\beta}$ was 0.1384. The corresponding approximate 95% and 99% confidence intervals were $[-2.370, -0.912]$ and $[-2.599, -0.683]$, respectively. When only time was partitioned into two subintervals, $\hat{\beta} = -1.587$. The corresponding approximate 95% and 99% confidence intervals were $[-2.369, -0.805]$ and $[-2.615, -0.560]$, respectively. All four partitions thus led to the same result.

Table 5. Recurrence times between the insertion of a catheter and the next infection of kidney patients.

Patient	Recurrence times		Event type *		
number	Y_1	Y_2	1st	2nd	Sex **
1	8	16	1	1	1
2	23	13	1	0	2
3	22	28	1	1	1
4	447	318	1	1	2
5	30	12	1	1	1
6	24	245	1	1	2
7	7	9	1	1	1
8	511	30	1	1	2
9	53	196	1	1	2
10	15	154	1	1	1
11	7	333	1	1	2
12	141	8	1	0	2
13	96	38	1	1	2
14	149	70	0	0	2
15	536	25	1	0	2
16	17	4	1	0	1
17	185	177	1	1	2
18	292	114	1	1	2
19	22	159	0	0	2
20	15	108	1	0	2
21	152	562	1	1	1
22	402	24	1	0	2
23	13	66	1	1	2
24	39	46	1	0	2
25	12	40	1	1	1
26	113	201	0	1	2
27	132	156	1	1	2
28	34	30	1	1	2
29	2	25	1	1	1
30	130	26	1	1	2
31	27	58	1	1	2
32	5	43	0	1	2
33	152	30	1	1	2
34	190	5	1	0	2
35	119	8	1	1	2
36	54	16	0	0	2
37	6	78	0	1	2
38	63	8	1	0	1

* Event type: 1=infection occurs, 0=censored;

** Sex: 1=male, 2=female.

Acknowledgement

We thank Professor Wei-Yann Tsai for his helpful discussions and comments. The research of Ying Chen was supported in part by the Research Foundation for Returnees Studied Abroad of Ministry of Education P.R.C. and supported in part by Shanghai Leading Academic Discipline Project, Project Number: B803. The research of Kani Chen was supported by Hong Kong RGC grants 600903 and 600706. The research of Zhiliang Ying was supported by grants from the US National Science Foundation and National Institutes of Health.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes, a large sample study. *Ann. Statist.* **10**, 1100-1120.

Bandeen-Roche, K. J. and Liang, K-Y. (1996). Modeling failure-time associations in data with multiple level of clustering. *Biometrika* **83**, 29-39.

Bandeen-Roche, K. J. and Liang, K-Y. (2002). Modeling multivariate failure-time associations in the presence of competing risk. *Biometrika* **89**, 299-314.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press.

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82**, 151-164.

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

Clayton, D. G. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Clayton, D. G. (1985). Multivariate generalization of the proportional hazards model (with discussion). *J. R. Statist. Soc. A* **148**, 82-117.

Clegg, J. X., Cai, J. and Sen, P. K. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* **55**, 805-812.

Cox, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

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

Fine, J., Yan, J. and Kosorok, M. R. (2004). Temporal process regression. *Biometrika* **91**, 683-703.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55**, 231-244.

Gray, R. J. and Li, Y. (2002). Optimal weight functions for marginal proportional hazards analysis of clustered failure time data. *Lifetime Data Anal.* **8**, 5-12.

Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc.* **55**, 698-707.

Horowitz, J. L. and Lee, S. (2004). Semiparametric estimation of a panel data proportional hazards model with fixed effects. *J. Econometrics* **119**, 155-198.

Hughes, M. D. (1995). Power considerations for clinical trials using multivariate time-to-event data. Manuscript.

Johnson, M. E. and Tenenbein, A. (1981). A bivariate distribution family with specified marginals. *J. Amer. Statist. Assoc.* **76**, 198-201.

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

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics* **47**, 461-466.

Oakes, D. (1992). Frailty models for multiple event-time data. In *Survival Analysis, State of the Art* (Edied by J. P. Klein and P. K. Goel), 371-380. Kluwer Academic Publishers.

Oakes, D. (1997). Model-based and/or marginal analysis for multiple event-time data? In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (Edited by D. Y. Lin and T. R. Fleming), 85-98. Springer-Verlag, New York.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analysis for recurrence failure times and time-dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

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

Prentice, R. L. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika* **84**, 349-363.

Qu, A., and Lindsay, B. G. (2003). Building adaptive estimating equations when inverse of covariance estimation is difficult. *J. Roy. Statist. Soc. Ser. B* **65**, 127-142.

Therneau, T. (1997). Extending the Cox model. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (Edited by D. Y. Lin and T. R. Fleming), 51-84. Springer-Verlag, New York.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling of marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Yang, Y. and Ying, Z. (2001). Marginal proportional hazards models for multiple event-time data. *Biometrika* **88**, 581-586.

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China.

E-mail: ychen@mail.shufe.edu.cn

Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong.

E-mail: makchen@ust.hk

Department of Statistics, Columbia University, New York, NY, 10027, USA.

E-mail: zying@stat.columbia.edu

(Received August 2008; accepted May 2009)