doi: 10.1111/j.1467-9469.2005.00487.x
 Vol 33: 1–23, 2006

# Rank Regression Analysis of Multivariate Failure Time Data Based on Marginal Linear Models

Z. JIN

*Department of Biostatistics, Columbia University*

D. Y. LIN

*Department of Biostatistics, University of North Carolina*

Z. YING

*Department of Statistics, Columbia University*

**ABSTRACT. Multivariate failure time data arises when each study subject can potentially experience several types of failures or recurrences of a certain phenomenon, or when failure times are sampled in clusters. We formulate the marginal distributions of such multivariate data with semi-parametric accelerated failure time models (i.e. linear regression models for log-transformed failure times with arbitrary error distributions) while leaving the dependence structures for related failure times completely unspecified. We develop rank-based monotone estimating functions for the regression parameters of these marginal models based on right-censored observations. The estimating equations can be easily solved via linear programming. The resultant estimators are consistent and asymptotically normal. The limiting covariance matrices can be readily estimated by a novel resampling approach, which does not involve non-parametric density estimation or evaluation of numerical derivatives. The proposed estimators represent consistent roots to the potentially non-monotone estimating equations based on weighted log-rank statistics. Simulation studies show that the new inference procedures perform well in small samples. Illustrations with real medical data are provided.**

*Key words:* accelerated failure time model, censoring, correlated data, linear programming, survival data, weighted log-rank statistics

## 1. Introduction

Multivariate failure time data are commonly encountered in scientific investigations because each study subject can potentially experience several events or because there exists natural or artificial clustering of study units such that the failure times within the same cluster are correlated. We refer to the former situation as multiple events data and the latter as clustered failure time data. An important special form of multiple events data are recurrent events data, which represents the repetitions of the same phenomenon. Statistical analysis of multivariate failure time data is complicated by right censoring as well as by the dependence of related failure times. Lin (1994) provided a review of Cox-type regression models for such data.

An important alternative to the Cox proportional hazards model is the accelerated failure time model (Kalbfleisch & Prentice, 2002, p. 44), which linearly regresses the logarithm of failure time on covariates. Rank estimation of the accelerated failure time model has been studied by Prentice (1978), Tsiatis (1990), Wei *et al.* (1990) and Lai & Ying (1991) among others for univariate failure time data, and by Lin & Wei (1992), Lee *et al.* (1993) and Lin *et al.* (1998) for multivariate failure time data. The rank estimators are derived from a class of weighted log-rank statistics. It is difficult to calculate the rank estimators because the estimating functions are step functions with multiple roots, some of which are inconsistent; identification of a consistent root can be very challenging in practice. A further difficulty lies in the variance–covariance estimation: the limiting covariance matrices of the rank estimators involve the unknown hazard function of the error term and are thus not amenable to numerical evaluations.

For univariate failure time data, some efforts have been made to alleviate the aforementioned difficulties. In particular, Jin *et al.* (2003) proposed a class of monotone estimating functions which approximates the weighted log-rank estimating functions around the true values of the regression parameters. The corresponding estimators are consistent and asymptotically normal with covariance matrices that can be readily estimated by a simple re-sampling technique. Both the parameter estimation and variance–covariance estimation can be performed via linear programming.

In this paper, we extend the work of Jin *et al.* (2003) to marginal accelerated failure time models for multivariate failure time data. We construct rank-based monotone estimating functions for three types of accelerated failure time models dealing with multiple events, recurrent events and clustered data. The resultant estimators are proven to be consistent and asymptotically normal. Furthermore, we develop a novel resampling approach which properly adjusts for the dependence of related failure times in the variance–covariance estimation. The proposed methods, like those of Jin *et al.* (2003), can be implemented efficiently through linear programming. Because of the intraclass correlation, the resampling scheme employed here is different from that of Jin *et al.* (2003) and entails considerable new technical challenges.

The rest of this paper is organized as follows. In sections 2–4, we present the models and corresponding inference procedures for multiple events data, recurrent events data and clustered failure time data respectively. In section 5, we report the results of our simulation studies. In section 6, we apply the proposed methods to two medical studies. Some concluding remarks are given in section 7. All the proofs are relegated to the appendix.

## 2. Multiple events data

### 2.1. Preliminaries

Consider a random sample of $n$ subjects, each of whom can potentially experience $K$ types of events or failures. For $i=1,\ldots,n$ and $k=1,\ldots,K$, let $T_{ki}$ be the time to the $k$th failure of the $i$th subject; let $C_{ki}$ be the corresponding censoring time, and $\mathbf{X}_{ki}$ be the corresponding $p_k \times 1$ vector of covariates. We assume that $(T_{1i},\ldots,T_{Ki})$ and $(C_{1i},\ldots,C_{Ki})$ are independent conditional on $(\mathbf{X}_{1i},\ldots,\mathbf{X}_{Ki})$. The data consists of $(\tilde{T}_{ki},\Delta_{ki},\mathbf{X}_{ki})$ $(k=1,\ldots,K; i=1,\ldots,n)$, where $\tilde{T}_{ki}=T_{ki}\wedge C_{ki}$ and $\Delta_{ki}=I(T_{ki}\le C_{ki})$. Here and in the sequel, $a\wedge b=\min(a,b)$ and $I(\cdot)$ is the indicator function.

We formulate the marginal distributions of the $K$ types of events with accelerated failure time models while leaving the dependence structures unspecified, i.e.

$$\log T_{ki}=\boldsymbol{\beta}_k'\mathbf{X}_{ki}+\epsilon_{ki}, \quad i=1,\ldots,n; \ k=1,\ldots,K,$$

where $\boldsymbol{\beta}_k \equiv (\beta_{1k},\ldots,\beta_{p_k k})'$ is a $p_k\times 1$ vector of unknown regression parameters, and $(\epsilon_{1i},\ldots,\epsilon_{Ki})(i=1,\ldots,n)$ are independent random vectors that are independent of the $\mathbf{X}_{ki}$ with a common, but completely unspecified, joint distribution.

Let

$$e_{ki}(\boldsymbol{\beta})=\log\tilde{T}_{ki}-\boldsymbol{\beta}'\mathbf{X}_{ki},$$

$$N_{ki}(\boldsymbol{\beta};t)=\Delta_{ki}I\{e_{ki}(\boldsymbol{\beta})\le t\}$$

and

$$\mathbf{S}_k^{(r)}(\boldsymbol{\beta};t)=n^{-1}\sum_{i=1}^{n} I\{e_{ki}(\boldsymbol{\beta})\geq t\}\mathbf{X}_{ki}^{r}(r=0,1).$$

The weighted log-rank estimating function for $\boldsymbol{\beta}_k$ is given by

$$\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})=\sum_{i=1}^{n}\Delta_{ki}\phi_k(\boldsymbol{\beta};e_{ki}(\boldsymbol{\beta}))\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};e_{ki}(\boldsymbol{\beta}))\},$$

or

$$\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\phi_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t),$$

where

$$\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)=\mathbf{S}_k^{(1)}(\boldsymbol{\beta};t)/S_k^{(0)}(\boldsymbol{\beta};t)$$

and $\phi_k$ is a weight function which satisfies condition 5 of Ying (1993, p. 90). The resultant estimator is denoted by $\widehat{\boldsymbol{\beta}}_{k,\phi_k}$. Note that the choices of 1, $S_k^{(0)}(\boldsymbol{\beta};t)$ and the Kaplan–Meier estimator based on $\{e_{ki}(\boldsymbol{\beta}),\Delta_{ki}\}$ $(i=1,\ldots,n)$ as $\phi_k(\boldsymbol{\beta};t)$ correspond to the log-rank, Gehan–Wilcoxon and Prentice–Wilcoxon statistics respectively.

Let

$$M_{ki}(\boldsymbol{\beta};t)=N_{ki}(\boldsymbol{\beta};t)-\int_0^t I\{e_{ki}(\boldsymbol{\beta})\geq u\}\lambda_k(u)\,\mathrm{d}u, \tag{1}$$

where $\lambda_k(\cdot)$ is the common hazard function of $\epsilon_{ki}\,(i=1,\ldots,n)$. Write $\mathbf{s}_k^{(r)}(\boldsymbol{\beta};t)=\lim_{n\to\infty}\mathbf{S}_k^{(r)}(\boldsymbol{\beta};t)$ $(r=0,1)$,

$$\overline{\mathbf{x}}_k(t)=\mathbf{s}_k^{(1)}(\boldsymbol{\beta}_k;t)/s_k^{(0)}(\boldsymbol{\beta}_k;t)$$

and

$$\phi_{0k}(t)=\lim_{n\to\infty}\phi_k(\boldsymbol{\beta}_k;t).$$

Define

$$\mathbf{A}_{k,\phi_k}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\phi_{0k}(t)\{\mathbf{X}_{ki}-\overline{\mathbf{x}}_k(t)\}^{\otimes 2}\left\{\frac{\mathrm{d}\log\lambda_k(t)}{\mathrm{d}t}\right\}\mathrm{d}N_{ki}(\boldsymbol{\beta}_k;t),$$

and

$$\mathbf{V}_{kl,\,\phi_k\phi_l}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathbf{u}_{ki,\phi_k}\mathbf{u}'_{li,\,\phi_l},$$

where $a^{\otimes 2}=aa'$ and

$$\mathbf{u}_{ki,\phi_k}=\int_{-\infty}^{\infty}\phi_{0k}(t)\{\mathbf{X}_{ki}-\overline{\mathbf{x}}_k(t)\}\,\mathrm{d}M_{ki}(\boldsymbol{\beta}_k;t). \tag{2}$$

Write $\mathbf{B}=(\boldsymbol{\beta}'_1,\ldots,\boldsymbol{\beta}'_K)'$ and $\widehat{\mathbf{B}}=(\widehat{\boldsymbol{\beta}}'_{1,\phi_1},\ldots,\widehat{\boldsymbol{\beta}}'_{K,\phi_K})'$. The random vector $n^{1/2}(\widehat{\mathbf{B}}-\mathbf{B})$ is asymptotically zero-mean normal with covariance matrix $\{\mathbf{A}_{k,\phi_k}^{-1}\mathbf{V}_{kl,\phi_k\phi_l}\mathbf{A}_{l,\phi_l}^{-1};k,l=1,\ldots,K\}$.

*2.2. Gehan weight function*

As mentioned earlier, the choice of $\phi_k(\boldsymbol{\beta}; t) = S_k^{(0)}(\boldsymbol{\beta}; t)$ corresponds to the Gehan (1965) weight function. In this case, $\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})$ can be expressed as

$$\mathbf{U}_{k,G}(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}(\mathbf{X}_{ki} - \mathbf{X}_{kj}) I\{e_{ki}(\boldsymbol{\beta}) \leq e_{kj}(\boldsymbol{\beta})\},$$

which is the gradient of the convex function

$$L_{k,G}(\boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}\{e_{ki}(\boldsymbol{\beta}) - e_{kj}(\boldsymbol{\beta})\}^{-},$$

where $a^{-} = I(a < 0)|a|$. Let $\widehat{\boldsymbol{\beta}}_{k,G}$ be a minimizer of $L_{k,G}(\boldsymbol{\beta})$. The minimization of $L_{k,G}(\boldsymbol{\beta})$ can be implemented by linear programming, and is equivalent to the minimization of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}|e_{ki}(\boldsymbol{\beta}) - e_{kj}(\boldsymbol{\beta})| + \left| Q - \boldsymbol{\beta}' \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}(\mathbf{X}_{kj} - \mathbf{X}_{ki}) \right|,$$

where $Q$ is any number which is greater than $\boldsymbol{\beta}' \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}(\mathbf{X}_{kj} - \mathbf{X}_{ki})$. This minimization can be implemented via an $L_1-$ minimization algorithm.

We shall approximate the joint distribution of the $\widehat{\boldsymbol{\beta}}_{k,G}$s by a resampling procedure. Let

$$L^{*}_{k,G}(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}\{e_{ki}(\boldsymbol{\beta}) - e_{kj}(\boldsymbol{\beta})\}^{-} Z_i Z_j, \quad k = 1, \ldots, K,$$

where $(Z_1, \ldots, Z_n)$ are independent positive random variables with $E(Z_i) = \mathrm{var}(Z_i) = 1$. It is important to note that the same set of $Z_i\,(i = 1, \ldots, n)$ is used in all the $K$ functions $L^{*}_{k,G}(\boldsymbol{\beta})$ $(k = 1, \ldots, K)$. Let $\widehat{\boldsymbol{\beta}}^{*}_{k,G}$ be a minimizer of $L^{*}_{k,G}(\boldsymbol{\beta})$ or a root of

$$\mathbf{U}^{*}_{k,G}(\boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ki}(\mathbf{X}_{ki} - \mathbf{X}_{kj}) I\{e_{ki}(\boldsymbol{\beta}) \leq e_{kj}(\boldsymbol{\beta})\} Z_i Z_j. \tag{3}$$

Again, $\widehat{\boldsymbol{\beta}}^{*}_{k,G}$ is obtained via linear programming. Write $\widehat{\mathbf{B}}^{*}_{G} = (\widehat{\boldsymbol{\beta}}^{*\prime}_{1,G}, \ldots, \widehat{\boldsymbol{\beta}}^{*\prime}_{K,G})'$ and $\widehat{\mathbf{B}}_{G} = (\widehat{\boldsymbol{\beta}}^{\prime}_{1,G}, \ldots, \widehat{\boldsymbol{\beta}}^{\prime}_{K,G})'$. We state below and prove in the appendix that the conditional distribution of $n^{1/2}(\widehat{\mathbf{B}}^{*}_{G} - \widehat{\mathbf{B}}_{G})$ given the data $(\tilde{T}_{ki}, \Delta_{ki}, \mathbf{X}_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$ can be used to approximate the distribution of $n^{1/2}(\widehat{\mathbf{B}}_{G} - \mathbf{B})$.

Conditional on the data $(\tilde{T}_{ki}, \Delta_{ki}, \mathbf{X}_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$, the only random elements in $L^{*}_{k,G}(\boldsymbol{\beta})$ $(k = 1, \ldots, K)$ are the $Z_i$s. To approximate the distribution of $\widehat{\mathbf{B}}_{G}$, we obtain a large number of realizations of $\widehat{\mathbf{B}}^{*}_{G}$ by repeatedly generating the random sample $(Z_1, \ldots, Z_n)$ while holding the data $(\tilde{T}_{ki}, \Delta_{ki}, \mathbf{X}_{ki})$ $(k = 1, \ldots, K; i = 1, \ldots, n)$ at their observed values. The covariance matrix of $\widehat{\mathbf{B}}_{G}$ can then be approximated by the empirical covariance matrix of $\widehat{\mathbf{B}}^{*}_{G}$.

To make our statements precise, we impose the following regularity conditions:

*Condition 1.* For $k = 1, \ldots, K$ and $i = 1, \ldots, n$, the Euclidean norms $\|\mathbf{X}_{ki}\|$ are bounded by a non-random constant.

*Condition 2.* Let $f_k(t)$ be the density function associated with $\lambda_k(t)$, $k = 1, \ldots, K$. Then $f_k(t)$ and $\mathrm{d}f_k(t)/\mathrm{d}t$ are bounded and $\int (\mathrm{d}\log f_k(s)/\mathrm{d}s)^2 f_k(s)\,\mathrm{d}s < \infty$.

*Condition 3.* The matrices $\mathbf{A}_{k,G}$ $(k = 1, \ldots, K)$ are non-singular, where $\mathbf{A}_{k,G}$ is $\mathbf{A}_{k,\phi_k}$ evaluated at $\phi_{0k} = s_k^{(0)}$.

*Remark 1.* Conditions 1 and 2 correspond to conditions 1 and 2 of Ying (1993) that are required to ensure the asymptotic linearity of the (weighted) log-rank estimating function. As indicated by Ying (1993), condition 1 may be relaxed to $\max_{k,i\le n}\|\mathbf{X}_{ki}\|=O(n^{\alpha})$ for any $\alpha>0$. It can be shown that all the commonly used error distributions satisfy condition 2. Condition 3 holds if for each $k$, the vector of covariates does not lie in a lower dimensional hyperplane, which is a minimum requirement for the identifiability of the regression parameters.

**Theorem 1**

*Under conditions 1–3, the estimator* $\widehat{\mathbf{B}}_G$ *is strongly consistent, and* $n^{1/2}(\widehat{\mathbf{B}}_G-\mathbf{B})$ *converges in distribution to a zero-mean multivariate normal random vector with covariance matrix* $\{\mathbf{A}_{k,G}^{-1}\mathbf{V}_{kl,G}\mathbf{A}_{l,G}^{-1};k,l=1,\dots,K\}$, *where* $\mathbf{V}_{kl,G}$ *is* $\mathbf{V}_{kl,\phi_k\phi_l}$ *evaluated at* $\phi_{0k}=s_k^{(0)}$ $(k=1,\dots,K)$. *Furthermore, the conditional distribution of* $n^{1/2}(\widehat{\mathbf{B}}_G^*-\widehat{\mathbf{B}}_G)$ *given the data* $(\tilde{T}_{ki},\Delta_{ki},\mathbf{X}_{ki})$ $(k=1,\dots,K;i=1,\dots,n)$ *converges almost surely to the same limiting distribution.*

The resampling scheme proposed here is different from that of Jin *et al.* (2003) even if $K=1$ in that each term in the summation of the perturbed function $L_{k,G}^*(\boldsymbol{\beta})$ is weighted by $Z_iZ_j$ rather than $Z_i$. This modification is required so as to properly account for the dependence of the multiple failure times within the same subject, and it creates significant technical challenges in the proofs.

As shown in the proof of theorem 1, $n^{-1}\mathbf{U}_{k,G}^*(\boldsymbol{\beta})$ has the same asymptotic slope as $n^{-1}\mathbf{U}_{k,G}(\boldsymbol{\beta})$ for each $k$, and the conditional joint distribution of $n^{-1/2}\{\mathbf{U}_{1,G}^*(\boldsymbol{\beta}_1),\dots,\mathbf{U}_{K,G}^*(\boldsymbol{\beta}_K)\}$ given the data $(\tilde{T}_{ki},\Delta_{ki},\mathbf{X}_{ki})$ $(k=1,\dots,K;i=1,\dots,n)$ converges to a zero-mean multivariate normal distribution whose covariance matrix is the limiting covariance matrix of $n^{-1/2}\{\mathbf{U}_{1,G}(\boldsymbol{\beta}_1),\dots,\mathbf{U}_{K,G}(\boldsymbol{\beta}_K)\}$. Thus, the conditional joint distribution of $n^{1/2}(\widehat{\mathbf{B}}_G^*-\widehat{\mathbf{B}}_G)$ given the data is the same in the limit as the joint distribution of $n^{1/2}(\widehat{\mathbf{B}}_G-\mathbf{B})$. If $Z_i$ instead of $Z_iZ_j$ were used in (3), then the conditional marginal distributions of $n^{-1/2}\{\mathbf{U}_{1,G}^*(\boldsymbol{\beta}_1),\dots,\mathbf{U}_{K,G}^*(\boldsymbol{\beta}_K)\}$ given the data would still be the same in the limit as the marginal distributions of $n^{-1/2}\{\mathbf{U}_{1,G}(\boldsymbol{\beta}_1),\dots,\mathbf{U}_{K,G}(\boldsymbol{\beta}_K)\}$, but the two *joint* distributions, specifically the two covariance matrices would be different.

### *2.3. General weight functions*

In general, $\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})$ is non-monotone. We consider the monotone modification of $\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})$:

$$\tilde{\mathbf{U}}_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_k)=\sum_{i=1}^{n}\Delta_{ki}\psi_k(\widehat{\boldsymbol{\beta}}_k;e_{ki}(\widehat{\boldsymbol{\beta}}_k))S_k^{(0)}(\boldsymbol{\beta};e_{ki}(\boldsymbol{\beta}))\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};e_{ki}(\boldsymbol{\beta}))\},$$

where

$$\psi_k(\boldsymbol{\beta};t)=\phi_k(\boldsymbol{\beta};t)/S_k^{(0)}(\boldsymbol{\beta};t)$$

and $\widehat{\boldsymbol{\beta}}_k$ is a preliminary consistent estimator of $\boldsymbol{\beta}_k$. Note that $\tilde{\mathbf{U}}_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_k)$ is monotone componentwise and is the gradient of the convex function

$$L_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_k)\equiv n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_k(\widehat{\boldsymbol{\beta}}_k;e_{ki}(\widehat{\boldsymbol{\beta}}_k))\Delta_{ki}\{e_{ki}(\boldsymbol{\beta})-e_{kj}(\boldsymbol{\beta})\}^{-},$$

which can again be minimized via linear programming. The minimization is carried out iteratively, i.e. $\widehat{\boldsymbol{\beta}}_{k(m)}=\arg\min_{\boldsymbol{\beta}}L_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{k(m-1)})$ $(m\ge 1)$, where $\widehat{\boldsymbol{\beta}}_{k(0)}=\widehat{\boldsymbol{\beta}}_{k,G}$. If the iterative algorithm converges as $m\to\infty$, then the limit satisfies the original estimating equation $\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})=\mathbf{0}$. For most commonly used weight functions, the algorithm converges stochastically in that, with a suitable choice of $m$ that depends on $n$, $\widehat{\boldsymbol{\beta}}_{k(m)}$ is asymptotically equivalent to the consistent roots of the original estimating equation $\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta})=\mathbf{0}$ (see Jin *et al.*, 2003). Whether the algorithm converges or not, $\widehat{\boldsymbol{\beta}}_{k(m)}$ is consistent and asymptotically normal.

To approximate the joint distribution of the $\widehat{\boldsymbol{\beta}}_{k(m)}$s, we again appeal to the resampling approach. Let $\widehat{\boldsymbol{\beta}}^*_{k(0)}=\widehat{\boldsymbol{\beta}}^*_{k,G}$ and $\widehat{\boldsymbol{\beta}}^*_{k(m)}=\arg\min_{\boldsymbol{\beta}} L^*_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}^*_{k(m-1)})$ $(m\geq 1)$, where

$$L^*_{k,\phi_k}(\boldsymbol{\beta};\mathbf{b})=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_k(\mathbf{b};e_{ki}(\mathbf{b}))\Delta_{ki}\{e_{ki}(\boldsymbol{\beta})-e_{kj}(\boldsymbol{\beta})\}^{-}Z_iZ_j.$$

Write $\widehat{\mathbf{B}}^*_{(m)}=(\widehat{\boldsymbol{\beta}}^{*\prime}_{1(m)},\ldots,\widehat{\boldsymbol{\beta}}^{*\prime}_{K(m)})'$ and $\widehat{\mathbf{B}}_{(m)}=(\widehat{\boldsymbol{\beta}}'_{1(m)},\ldots,\widehat{\boldsymbol{\beta}}'_{K(m)})'$. We state below and prove in the appendix that, for any $m$, the conditional distribution of $n^{1/2}(\widehat{\mathbf{B}}^*_{(m)}-\widehat{\mathbf{B}}_{(m)})$ given the data is asymptotically equivalent to the limiting distribution of $n^{1/2}(\widehat{\mathbf{B}}_{(m)}-\mathbf{B})$.

We impose two additional regularity conditions:

*Condition 4.* For each $k=1,\ldots,K$, both $\mathbf{A}_{k,\phi_k}$ and $(\mathbf{A}_{k,\phi_k}+\mathbf{D}_{k,\phi_k})$ are non-singular, where

$$\mathbf{D}_{k,\phi_k}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\dot{\psi}_{0k}(t)s_k^{(0)}(\boldsymbol{\beta}_k;t)\{\mathbf{X}_{ki}-\overline{\mathbf{x}}_k(t)\}^{\otimes 2}\,\mathrm{d}N_{ki}(\boldsymbol{\beta}_k;t),$$

and $\dot{\psi}_{0k}(t)$ is the derivative of $\psi_{0k}(t)\equiv\lim_{n\to\infty}\psi_k(\boldsymbol{\beta}_k;t)$.

*Condition 5.* For each $k=1,\ldots,K$ and for any $\boldsymbol{\beta}_n$ and $\eta_n$ such that $\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_k\|+|\eta_n|=o(n^{-\epsilon})$ almost surely for some $\epsilon>0, \psi_k(\boldsymbol{\beta}_n;t)=\psi_k(\boldsymbol{\beta}_k;t)+o(1)$ and $\psi_k(\boldsymbol{\beta}_n;t+\eta_n)=\psi_k(\boldsymbol{\beta}_n;t)+\dot{\psi}_{0k}(t)\eta_n+o(n^{-1/2}+\eta_n)$, both uniformly in $t$.

**Theorem 2**

*Suppose that conditions 1–5 hold. For each m, the estimator* $\widehat{\mathbf{B}}_{(m)}$ *is strongly consistent, and* $n^{1/2}(\widehat{\mathbf{B}}_{(m)}-\mathbf{B})$ *converges to a zero-mean multivariate normal distribution. Furthermore, the conditional distribution of* $n^{1/2}(\widehat{\mathbf{B}}^*_{(m)}-\widehat{\mathbf{B}}_{(m)})$ *given the data* $(\tilde{T}_{ki},\Delta_{ki},\mathbf{X}_{ki})$ $(k=1,\ldots,K;i=1,\ldots,n)$ *converges almost surely to the same limiting distribution.*

For notational simplicity, we shall drop the subscript $(m)$ in $\widehat{\mathbf{B}}^*_{(m)}$ and $\widehat{\mathbf{B}}_{(m)}$. To approximate the distribution of $\widehat{\mathbf{B}}$, we obtain a large number of realizations of $\widehat{\mathbf{B}}^*$ by repeatedly generating the random sample $(Z_1,\ldots,Z_n)$ while fixing the data $(\tilde{T}_{ki},\Delta_{ki},\mathbf{X}_{ki})$ $(k=1,\ldots,K;i=1,\ldots,n)$ at their observed values. The covariance matrix of $\widehat{\mathbf{B}}$ can then be approximated by the empirical covariance matrix of $\widehat{\mathbf{B}}^*$, denoted by $\widehat{\mathbf{V}}$.

The above results enable one to carry out simultaneous inference on $\mathbf{B}$. Suppose, for example, one is interested in the effects $\eta_k\equiv\beta_{1k}$ $(k=1,\ldots,K)$ of a particular kind of covariate on the $K$ event times. Let $\widehat{\mathbf{V}}_\eta$ be the part of $\widehat{\mathbf{V}}$ corresponding to the covariance matrix of $(\widehat{\eta}_1,\ldots,\widehat{\eta}_K)'$, where $\widehat{\eta}_k=\widehat{\beta}_{1k}$. Then the null hypothesis $H_0:\eta_1=\eta_2=\cdots=\eta_K=0$ can be tested by using the quadratic form $(\widehat{\eta}_1,\ldots,\widehat{\eta}_K)\widehat{\mathbf{V}}_\eta^{-1}(\widehat{\eta}_1,\ldots,\widehat{\eta}_K)'$. One can also determine which of the individual hypotheses $\eta_k=0$ $(k=1,\ldots,K)$ should be rejected by using the sequential multiple testing procedures discussed in Wei *et al.* (1989). Under the restriction that $\eta_1=\eta_2=\cdots=\eta_K=\eta$, the optimal linear estimator $\widehat{\eta}\equiv\sum_{k=1}^{K}c_k\widehat{\eta}_k$, where $(c_1,\ldots,c_K)'=(\mathbf{1}'\widehat{\mathbf{V}}_\eta^{-1}\mathbf{1})^{-1}\widehat{\mathbf{V}}_\eta^{-1}\mathbf{1}$ and $\mathbf{1}=(1,\ldots,1)'$, has the smallest asymptotic variance among all linear estimators for $\eta$.

## 3. Recurrent events data

### *3.1. Preliminaries*

Suppose that we have a random sample of $n$ subjects. For $i=1,\dots,n$ and $k=1,2,\dots,$ let $T_{ki}$ be the time to the $k$th recurrent event on the $i$th subject; let $C_i$ and $\mathbf{X}_i$ be the censoring time and the $p\times 1$ vector of covariates for the $i$th subject. Assume that $C_i$ is independent of $T_{ki}$ $(k=1,2,\dots)$ conditional on $\mathbf{X}_i$. Let

$$N_i^*(t)=\sum_{k=1}^{\infty} I(T_{ki}\le t).$$

We specify the following accelerated time model for the mean frequency function:

$$E\{N_i^*(t)|\mathbf{X}_i\}=\mu_0(t\,\mathrm{e}^{-\boldsymbol{\beta}_0'X_i}), \tag{4}$$

where $\boldsymbol{\beta}_0$ is a $p\times 1$ vector of regression parameters, and $\mu_0(\cdot)$ is an unspecified baseline mean function. The weighted log-rank estimating function for $\boldsymbol{\beta}_0$ takes the form

$$\mathbf{U}_\phi(\boldsymbol{\beta})=\sum_{i=1}^{n}\sum_{k=1}^{\infty} I(T_{ki}\le C_i)\phi(\boldsymbol{\beta};T_{ki}\,\mathrm{e}^{-\boldsymbol{\beta}'\mathbf{X}_i})\{\mathbf{X}_i-\overline{\mathbf{X}}(\boldsymbol{\beta};T_{ki}\,\mathrm{e}^{-\boldsymbol{\beta}'\mathbf{X}_i})\}, \tag{5}$$

where

$$\overline{\mathbf{X}}(\boldsymbol{\beta};t)=\frac{\mathbf{S}^{(1)}(\boldsymbol{\beta};t)}{S^{(0)}(\boldsymbol{\beta};t)},\quad \mathbf{S}^{(r)}(\boldsymbol{\beta};t)=n^{-1}\sum_{j=1}^{n} I(C_j\,\mathrm{e}^{-\boldsymbol{\beta}'\mathbf{X}_j}\ge t)\mathbf{X}_j^r\,(r=0,1)$$

and $\phi$ is a weight function. The resultant estimator $\widehat{\boldsymbol{\beta}}_\phi$ is consistent and asymptotically normal.

### *3.2. Gehan weight function*

Lin *et al.* (1998) noted that, for the Gehan weight function, $\mathbf{U}_\phi(\boldsymbol{\beta})$ reduces to

$$\mathbf{U}_G(\boldsymbol{\beta})=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} I(T_{ki}\le C_i)(\mathbf{X}_i-\mathbf{X}_j)I\{\log T_{ki}-\log C_j\le \boldsymbol{\beta}'(\mathbf{X}_i-\mathbf{X}_j)\}.$$

Thus, the corresponding estimator $\widehat{\boldsymbol{\beta}}_G$ can be obtained by minimizing the convex function

$$L_G(\boldsymbol{\beta})=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} I(T_{ki}\le C_i)\{\log T_{ki}-\log C_j-\boldsymbol{\beta}'(\mathbf{X}_i-\mathbf{X}_j)\}^-$$

via linear programming. Define

$$L_G^*(\boldsymbol{\beta})=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty} I(T_{ki}\le C_i)\{\log T_{ki}-\log C_j-\boldsymbol{\beta}'(\mathbf{X}_i-\mathbf{X}_j)\}^- Z_iZ_j,$$

where $(Z_1,\dots,Z_n)$ are the same as in section 2. Denote a minimizer of $L_G^*(\boldsymbol{\beta})$ by $\widehat{\boldsymbol{\beta}}_G^*$, which again is obtained via linear programming.

Let

$$N_i(\boldsymbol{\beta};t)=N_i^*(t\,\mathrm{e}^{\boldsymbol{\beta}'\mathbf{X}_i}\wedge C_i)$$

and

$$M_i(\boldsymbol{\beta};t)=N_i(\boldsymbol{\beta};t)-\int_0^t I(C_i\,\mathrm{e}^{-\boldsymbol{\beta}'\mathbf{X}_i}\ge u)\,\mathrm{d}\mu_0(u).$$

Also, let

$$s^{(r)}(\boldsymbol{\beta};t)=\lim_{n\to\infty}\mathbf{S}^{(r)}(\boldsymbol{\beta};t)\,(r=0,1)$$

and

$$\bar{\mathbf{x}}(t) = \mathbf{s}^{(1)}(\boldsymbol{\beta}_0; t)/s^{(0)}(\boldsymbol{\beta}_0; t).$$

Define

$$\mathbf{A}_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^\infty \phi_0(t) I(C_i \, \mathrm{e}^{-\boldsymbol{\beta}_0' \mathbf{X}_i} \geq t) \{\mathbf{X}_i - \bar{\mathbf{x}}(t)\}^{\otimes 2} \, \mathrm{d}\{\dot{\mu}_0(t) t\},$$

and

$$\mathbf{V}_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{u}_{i,\phi}^{\otimes 2},$$

where

$$\mathbf{u}_{i,\phi} = \int_0^\infty \phi_0(t) \{\mathbf{X}_i - \bar{\mathbf{x}}(t)\} \, \mathrm{d}M_i(\boldsymbol{\beta}_0; t),$$

$$\phi_0(t) = \lim_{n\to\infty} \phi(\boldsymbol{\beta}_0; t),$$

and

$$\dot{\mu}_0(t) = \mathrm{d}\mu_0(t)/\mathrm{d}t.$$

We impose the following conditions:

*Condition 6.* For all $i$, $\|\mathbf{X}_i\| + C_i + N_i^*(C_i)$ are bounded by a non-random constant.

*Condition 7.* The function $\mu_0$ is continuously differentiable.

*Condition 8.* The matrix $\mathbf{A}_G$ is non-singular, where $\mathbf{A}_G$ is $\mathbf{A}_\phi$ evaluated at $\phi_0 = s^{(0)}$.

**Theorem 3**

*Under conditions 6–8, the estimator $\widehat{\boldsymbol{\beta}}_G$ is strongly consistent, and $n^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\mathbf{A}_G^{-1}\mathbf{V}_G\mathbf{A}_G^{-1}$, where $\mathbf{V}_G$ is $\mathbf{V}_\phi$ evaluated at $\phi_0 = s^{(0)}$. Furthermore, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G^* - \widehat{\boldsymbol{\beta}}_G)$ given the data $(C_i, T_{ki}, \mathbf{X}_i)$ $(T_{ki} \leq C_i;\ i = 1, \ldots, n)$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$.*

### *3.3. General weight functions*

To approximate $\widehat{\boldsymbol{\beta}}_\phi$ and its covariance matrix, we define

$$L_\phi(\boldsymbol{\beta}; \mathbf{b}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} \psi(\mathbf{b}; T_{ki} \, \mathrm{e}^{-\mathbf{b}'\mathbf{X}_i}) I(T_{ki} \leq C_i) \{\log T_{ki} - \log C_j - \boldsymbol{\beta}'(\mathbf{X}_i - \mathbf{X}_j)\}^-,$$

$$L_\phi^*(\boldsymbol{\beta}; \mathbf{b}) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{\infty} \psi(\mathbf{b}; T_{ki} \, \mathrm{e}^{-\mathbf{b}'\mathbf{X}_i}) I(T_{ki} \leq C_i) \{\log T_{ki} - \log C_j - \boldsymbol{\beta}'(\mathbf{X}_i - \mathbf{X}_j)\}^- Z_i Z_j,$$

where $\psi(\boldsymbol{\beta}; t) = \phi(\boldsymbol{\beta}; t)/S^{(0)}(\boldsymbol{\beta}; t)$. For $m \geq 1$, let $\widehat{\boldsymbol{\beta}}_{(m)} = \arg\min_{\boldsymbol{\beta}} L_\phi(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}_{(m-1)})$, and $\widehat{\boldsymbol{\beta}}_{(m)}^* = \arg\min_{\boldsymbol{\beta}} L_\phi^*(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}_{(m-1)}^*)$, where $\widehat{\boldsymbol{\beta}}_{(0)} = \widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}_{(0)}^* = \widehat{\boldsymbol{\beta}}_G^*$. We impose the following conditions.

*Condition 9.* Both $\mathbf{A}_\phi$ and $(\mathbf{A}_\phi + \mathbf{D}_\phi)$ are non-singular, where

$$\mathbf{D}_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_0^\infty \dot{\psi}_0(t) t s^{(0)}(\boldsymbol{\beta}_0; t)\{\mathbf{X}_i - \overline{\mathbf{x}}(t)\}^{\otimes 2}\, \mathrm{d}N_i(\boldsymbol{\beta}_0; t),$$

and $\dot{\psi}_0(t)$ is the derivative of $\psi_0(t) \equiv \lim_{n\to\infty} \psi(\boldsymbol{\beta}_0; t)$.

*Condition 10.* For any $\boldsymbol{\beta}_n$ and $\eta_n$ such that $\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_0\| + |\eta_n| = o(n^{-\epsilon})$ almost surely for some $\epsilon > 0, \psi(\boldsymbol{\beta}_n; t) = \psi(\boldsymbol{\beta}_0; t) + o(1)$ and $\psi(\boldsymbol{\beta}_n; t(1+\eta_n)) = \psi(\boldsymbol{\beta}_n; t) + \dot{\psi}_0(t)\eta_n + o(n^{-1/2} + \eta_n)$, both uniformly in $t \le \tau$, where $\tau = \sup\{t : \Pr(C\,\mathrm{e}^{-\boldsymbol{\beta}_0'\mathbf{X}} \ge t) > 0\}$.

**Theorem 4**

*Suppose that conditions 6–10 are satisfied. For each m, the estimator $\widehat{\boldsymbol{\beta}}_{(m)}$ is strongly consistent, and $n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0)$ converges to a zero-mean multivariate normal distribution. Furthermore, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^*_{(m)} - \widehat{\boldsymbol{\beta}}_{(m)})$ given the data $(C_i, T_{ki}, \mathbf{X}_i)$ $(T_{ki} \le C_i; i = 1, \ldots, n)$ converges almost surely to the same limiting distribution.*

## 4. Clustered failure time data

### *4.1. Preliminaries*

Suppose that we have a random sample of $n$ clusters and there are $K_i$ members in the $i$th cluster. Let $T_{ik}$ and $C_{ik}$ be the failure time and censoring time for the $k$th member of the $i$th cluster, and let $\mathbf{X}_{ik}$ be the corresponding $p \times 1$ vector of covariates. We assume that $(T_{i1}, \ldots, T_{iK_i})$ and $(C_{i1}, \ldots, C_{iK_i})$ are independent conditional on $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iK_i})$. The data consist of $(\tilde{T}_{ik}, \Delta_{ik}, \mathbf{X}_{ik})$ $(k = 1, \ldots, K_i; i = 1, \ldots, n)$, where $\tilde{T}_{ik} = T_{ik} \wedge C_{ik}$ and $\Delta_{ik} = I(T_{ik} \le C_{ik})$.

We specify that the marginal distributions of the $T_{ik}$ satisfy the accelerated failure time model:

$$\log T_{ik} = \boldsymbol{\beta}_0'\mathbf{X}_{ik} + \epsilon_{ik}, \quad k = 1, \ldots, K_i; \;\; i = 1, \ldots, n,$$

where $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters, and $(\epsilon_{i1}, \ldots, \epsilon_{iK_i})$ $(i = 1, \ldots, n)$ are independent random vectors. For each $i$, the error terms $\epsilon_{i1}, \ldots, \epsilon_{iK_i}$ are potentially correlated, but are assumed to be exchangeable with a common marginal distribution; for any $i$ and $j$, and $K \le K_i \wedge K_j$, the vectors $(\epsilon_{i1}, \ldots, \epsilon_{iK})'$ and $(\epsilon_{j1}, \ldots, \epsilon_{jK})'$ have the same distribution.

Let $e_{ik}(\boldsymbol{\beta}) = \log \tilde{T}_{ik} - \boldsymbol{\beta}'\mathbf{X}_{ik}$ and $\mathbf{S}^{(r)}(\boldsymbol{\beta}; t) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i} I\{e_{ik}(\boldsymbol{\beta}) \ge t\}\mathbf{X}_{ik}^{r} (r = 0, 1)$. Under the independence working assumption, the weighted log-rank estimating function takes the form

$$\mathbf{U}_\phi(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K_i} \Delta_{ik}\phi\Big(\boldsymbol{\beta}; e_{ik}(\boldsymbol{\beta})\Big)\Big\{\mathbf{X}_{ik} - \overline{\mathbf{X}}\Big(\boldsymbol{\beta}; e_{ik}(\boldsymbol{\beta})\Big)\Big\}, \tag{6}$$

where $\overline{\mathbf{X}}(\boldsymbol{\beta}; t) = \mathbf{S}^{(1)}(\boldsymbol{\beta}; t)/S^{(0)}(\boldsymbol{\beta}; t)$ and $\phi$ is a weight function. Denote the estimator by $\widehat{\boldsymbol{\beta}}_\phi$.

### *4.2. Gehan weight function*

For $\phi(\boldsymbol{\beta}; t) = S^{(0)}(\boldsymbol{\beta}; t)$, we can express $\mathbf{U}_\phi(\boldsymbol{\beta})$ as

$$\mathbf{U}_G(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{j=1}^{n}\sum_{l=1}^{K_j} \Delta_{ik}(\mathbf{X}_{ik} - \mathbf{X}_{jl}) I\{e_{ik}(\boldsymbol{\beta}) \le e_{jl}(\boldsymbol{\beta})\},$$

which is the gradient of

$$L_G(\boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \sum_{l=1}^{K_j} \Delta_{ik} \{e_{ik}(\boldsymbol{\beta}) - e_{jl}(\boldsymbol{\beta})\}^{-}.$$

Let $\widehat{\boldsymbol{\beta}}_G$ be a minimizer of $L_G(\boldsymbol{\beta})$, which can again be obtained by linear programming. Define

$$L_G^*(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \sum_{l=1}^{K_j} \Delta_{ik} \{e_{ik}(\boldsymbol{\beta}) - e_{jl}(\boldsymbol{\beta})\}^{-} Z_i Z_j,$$

where $(Z_1, \ldots, Z_n)$ are defined in section 2. Let $\widehat{\boldsymbol{\beta}}_G^*$ be a minimizer of $L_G^*(\boldsymbol{\beta})$.

Define $N_{ik}(\boldsymbol{\beta}; t) = \Delta_{ik} I\{e_{ik}(\boldsymbol{\beta}) \le t\}$ and $M_{ik}(\boldsymbol{\beta}; t) = N_{ik}(\boldsymbol{\beta}; t) - \int_{-\infty}^{t} I\{e_{ik}(\boldsymbol{\beta}) \ge u\} \lambda_0(u)\, \mathrm{d}u$, where $\lambda_0(\cdot)$ is the common hazard function of the $\epsilon_{ik}$s. Also, define

$$\mathbf{A}_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \int_{-\infty}^{\infty} \phi_0(t) \{\mathbf{X}_{ik} - \overline{\mathbf{x}}(t)\}^{\otimes 2} \left\{ \frac{\mathrm{d} \log \lambda_0(t)}{\mathrm{d}t} \right\} \mathrm{d}N_{ik}(\boldsymbol{\beta}_0; t),$$

and

$$\mathbf{V}_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left( \sum_{k=1}^{K_i} \mathbf{u}_{ik,\phi} \right)^{\otimes 2},$$

where

$$\mathbf{u}_{ik,\phi} = \int_{-\infty}^{\infty} \phi_0(t) \{\mathbf{X}_{ik} - \overline{\mathbf{x}}(t)\}\, \mathrm{d}M_{ik}(\boldsymbol{\beta}_0; t),$$

$$\phi_0(t) = \lim_{n\to\infty} \phi(\boldsymbol{\beta}_0; t),$$

$$\overline{\mathbf{x}}(t) = \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0; t)}{s^{(0)}(\boldsymbol{\beta}_0; t)},$$

and $\mathbf{s}^{(r)}(\boldsymbol{\beta}; t) = \lim_{n\to\infty} \mathbf{S}^{(r)}(\boldsymbol{\beta}; t) (r = 0, 1)$. We impose the following regularity conditions:

*Condition 11.* For all $i$, $\sum_{k=1}^{K_i} \|\mathbf{X}_{ik}\| + K_i$ are bounded by a nonrandom constant.

*Condition 12.* Let $f$ be the density function associated with $\lambda_0$. Then $f(t)$ and $\mathrm{d}f(t)/\mathrm{d}t$ are bounded, and $\int \{\mathrm{d} \log f(t)/\mathrm{d}t\}^2 f(t) \mathrm{d}t < \infty$.

*Condition 13.* The matrix $\mathbf{A}_G$ is non-singular, where $\mathbf{A}_G$ is $\mathbf{A}_\phi$ evaluated at $\phi_0 = s^{(0)}$.

**Theorem 5**

*Under conditions 11–13, the estimator $\widehat{\boldsymbol{\beta}}_G$ is stongly consistent, and $n^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\mathbf{A}_G^{-1} \mathbf{V}_G \mathbf{A}_G^{-1}$, where $\mathbf{V}_G$ is $\mathbf{V}_\phi$ evaluated at $\phi_0 = s^{(0)}$. Furthermore, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G^* - \widehat{\boldsymbol{\beta}}_G)$ given the data $(\tilde{T}_{ik}, \Delta_{ik}, \mathbf{X}_{ik})$ $(k = 1, \ldots, K_i;\ i = 1, \ldots, n)$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$.*

*4.3. General weight functions*

We consider

$$L_{\phi}(\boldsymbol{\beta};\mathbf{b})\equiv n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{j=1}^{n}\sum_{l=1}^{K_j}\psi\Big(\mathbf{b};e_{ik}(\mathbf{b})\Big)\Delta_{ik}\{e_{ik}(\boldsymbol{\beta})-e_{jl}(\boldsymbol{\beta})\}^{-},$$

$$L^{*}_{\phi}(\boldsymbol{\beta};\mathbf{b})\equiv n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{j=1}^{n}\sum_{l=1}^{K_j}\psi\Big(\mathbf{b};e_{ik}(\mathbf{b})\Big)\Delta_{ik}\{e_{ik}(\boldsymbol{\beta})-e_{jl}(\boldsymbol{\beta})\}^{-}Z_iZ_j,$$

where $\psi(\boldsymbol{\beta};t)=\phi(\boldsymbol{\beta};t)/S^{(0)}(\boldsymbol{\beta};t)$. For $m\geq 1$, let $\widehat{\boldsymbol{\beta}}_{(m)}=\arg\min_{\boldsymbol{\beta}}L_{\phi}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{(m-1)})$ and $\widehat{\boldsymbol{\beta}}^{*}_{(m)}=\arg\min_{\boldsymbol{\beta}}L^{*}_{\phi}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}^{*}_{(m-1)})$, where $\widehat{\boldsymbol{\beta}}_{(0)}=\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}^{*}_{(0)}=\widehat{\boldsymbol{\beta}}^{*}_G$. We impose two additional conditions:

*Condition 14.* Both $\mathbf{A}_{\phi}$ and $(\mathbf{A}_{\phi}+\mathbf{D}_{\phi})$ are non-singular, where

$$\mathbf{D}_{\phi}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{-\infty}^{\infty}\dot{\psi}_0(t)s^{(0)}(\boldsymbol{\beta}_0;t)\{\mathbf{X}_{ik}-\overline{\mathbf{x}}(t)\}^{\otimes 2}\,\mathrm{d}N_{ik}(\boldsymbol{\beta}_0;t),$$

and $\dot{\psi}_0(t)$ is the derivative of $\psi_0(t)\equiv\lim_{n\to\infty}\psi(\boldsymbol{\beta}_0;t)$.

*Condition 15.* For any $\boldsymbol{\beta}_n$ and $\eta_n$ such that $\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|+|\eta_n|=o(n^{-\epsilon})$ almost surely for some $\epsilon>0$, $\psi(\boldsymbol{\beta}_n;t)=\psi(\boldsymbol{\beta}_0;t)+o(1)$ and $\psi(\boldsymbol{\beta}_n;t+\eta_n)=\psi(\boldsymbol{\beta}_n;t)+\dot{\psi}_0(t)\eta_n+o(n^{-1/2}+\eta_n)$, both uniformly in $t$.

**Theorem 6**

*Suppose that conditions 11–15 are satisfied. For each m, the estimator $\widehat{\boldsymbol{\beta}}_{(m)}$ is strongly consistent and $n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}_0)$ converges to a zero-mean multivariate normal distribution. Furthermore, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^{*}_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)})$ given the data $(\tilde{T}_{ik},\Delta_{ik},\mathbf{X}_{ik})$ $(k=1,\ldots,K_i;$ $i=1,\ldots,n)$ converges almost surely to the same limiting distribution.*

## 5. Numerical studies

We carried out extensive simulation studies to evaluate the small-sample properties of the methods developed in sections 2–4. We focused on the Gehan and log-rank weight functions. The (approximate) log-rank estimates were obtained with three iterations. The differences between the estimates with three iterations and those at convergence are generally negligible.

For multiple events and clustered data, two failure times $T_1$ and $T_2$ were generated from Gumbel (1960) bivariate distribution:

$$F(t_1,t_2)=F_1(t_1)F_2(t_2)[1+\theta\{1-F_1(t_1)\}\{1-F_2(t_2)\}],$$

where $-1\leq\theta\leq 1$. The correlation between $T_1$ and $T_2$ is $\theta/4$. The two marginal distributions $F_k(t_k)$ $(k=1,2)$ were exponential with hazard rates $\lambda_k=\mathrm{e}^{\beta_1X_{1k}+\beta_2X_{2k}}$, where $X_{1k}$ $(k=1,2)$ were Bernoulli with 0.5 success probability and $X_{2k}$ $(k=1,2)$ were independent standard normal truncated at $\pm 2$. For multiple events, $T_1$ and $T_2$ shared the same set of covariates, i.e. $X_{11}=X_{12}$ and $X_{21}=X_{22}$; for clustered data, the covariates were generated independently. The censoring times were generated from the uniform $(0, \tau)$ distribution, where $\tau$ was chosen to yield a desired level of censoring.

For recurrent events, the covariates were generated in the same manner as in the case of multiple events. The gap times between successive events were generated from the aforementioned Gumbel's bivariate exponential distribution. The resultant recurrent event process is Poisson under $\theta=0$ and non-Poisson under $\theta \neq 0$. The follow-up time was an independent uniform $(0, 2.5)$ random variable, which on average yielded approximately 2.60 and 2.86 events per subject for the Poisson and non-Poisson cases respectively.

Tables 1 and 2 summarizes the results on the estimation of $\beta_1$ when $\beta_1=1$ and $\beta_2=0.5$. The results for $\beta_2$ are similar and thus omitted. Each entry in the table was based on 1000 simulated data sets. For each data set, we approximated the limiting distribution of the para-

Table 1. *Simulation results for multiple events and clustered data*

| | | | | Multiple events | | | | Clustered data | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | $n$ | Censoring (%) | Weight | Bias | SE | SEE | CP | Bias | SE | SEE | CP |
| 0 | 50 | 0 | Gehan | 0.001 | 0.245 | 0.238 | 0.945 | 0.002 | 0.241 | 0.233 | 0.940 |
| | | | Log-rank | 0.003 | 0.218 | 0.213 | 0.937 | 0.004 | 0.211 | 0.206 | 0.934 |
| | | 25 | Gehan | 0.012 | 0.364 | 0.360 | 0.937 | −0.041 | 0.353 | 0.357 | 0.951 |
| | | | Log-rank | 0.021 | 0.334 | 0.343 | 0.948 | −0.042 | 0.327 | 0.333 | 0.957 |
| | | 50 | Gehan | 0.019 | 0.518 | 0.481 | 0.922 | −0.067 | 0.550 | 0.534 | 0.944 |
| | | | Log-rank | 0.016 | 0.487 | 0.491 | 0.952 | −0.073 | 0.526 | 0.515 | 0.952 |
| | 100 | 0 | Gehan | −0.007 | 0.165 | 0.166 | 0.957 | −0.002 | 0.160 | 0.164 | 0.960 |
| | | | Log-rank | −0.004 | 0.147 | 0.146 | 0.949 | −0.002 | 0.144 | 0.145 | 0.946 |
| | | 25 | Gehan | −0.003 | 0.246 | 0.247 | 0.950 | 0.001 | 0.245 | 0.246 | 0.947 |
| | | | Log-rank | 0.002 | 0.223 | 0.229 | 0.947 | −0.004 | 0.223 | 0.226 | 0.954 |
| | | 50 | Gehan | 0.005 | 0.351 | 0.355 | 0.949 | −0.005 | 0.348 | 0.357 | 0.953 |
| | | | Log-rank | 0.008 | 0.329 | 0.344 | 0.956 | −0.011 | 0.319 | 0.336 | 0.956 |
| 1 | 50 | 0 | Gehan | −0.007 | 0.289 | 0.273 | 0.930 | 0.010 | 0.234 | 0.233 | 0.944 |
| | | | Log-rank | −0.011 | 0.249 | 0.239 | 0.937 | 0.011 | 0.202 | 0.207 | 0.953 |
| | | 25 | Gehan | 0.009 | 0.408 | 0.397 | 0.939 | −0.028 | 0.354 | 0.352 | 0.953 |
| | | | Log-rank | 0.017 | 0.373 | 0.378 | 0.942 | −0.027 | 0.328 | 0.328 | 0.950 |
| | | 50 | Gehan | 0.031 | 0.560 | 0.528 | 0.923 | −0.064 | 0.542 | 0.526 | 0.936 |
| | | | Log-rank | 0.036 | 0.521 | 0.541 | 0.949 | −0.067 | 0.506 | 0.509 | 0.947 |
| | 100 | 0 | Gehan | 0.007 | 0.186 | 0.190 | 0.951 | 0.005 | 0.165 | 0.164 | 0.956 |
| | | | Log-rank | 0.006 | 0.165 | 0.163 | 0.946 | 0.004 | 0.143 | 0.143 | 0.950 |
| | | 25 | Gehan | −0.002 | 0.268 | 0.276 | 0.952 | −0.007 | 0.249 | 0.244 | 0.952 |
| | | | Log-rank | 0.001 | 0.250 | 0.257 | 0.960 | −0.005 | 0.230 | 0.226 | 0.940 |
| | | 50 | Gehan | 0.007 | 0.387 | 0.379 | 0.937 | −0.007 | 0.370 | 0.356 | 0.941 |
| | | | Log-rank | 0.010 | 0.368 | 0.370 | 0.953 | −0.012 | 0.336 | 0.336 | 0.951 |

Bias and SE represents the bias and standard error of the estimator, SEE represents the mean of the standard error estimator and CP represents the coverage probability of the 95% Wald-type confidence interval. The results for multiple events pertain to the optimal linear estimator.

Table 2. *Simulation results for recurrent events data*

| $\theta$ | $n$ | Weight | Bias | SE | SEE | CP |
|---|---|---|---|---|---|---|
| 0 | 50 | Gehan | −0.022 | 0.249 | 0.235 | 0.935 |
| | | Log-rank | −0.019 | 0.237 | 0.223 | 0.932 |
| | 100 | Gehan | −0.009 | 0.165 | 0.165 | 0.941 |
| | | Log-rank | −0.009 | 0.157 | 0.156 | 0.941 |
| 1 | 50 | Gehan | −0.004 | 0.269 | 0.261 | 0.946 |
| | | Log-rank | −0.008 | 0.265 | 0.253 | 0.938 |
| | 100 | Gehan | 0.007 | 0.188 | 0.185 | 0.942 |
| | | Log-rank | 0.004 | 0.178 | 0.179 | 0.950 |

For explanation see Table 1.

meter estimator using 1000 samples of $(Z_1, \ldots, Z_n)$, where the $Z_i$s are standard exponential random variables. The simulation results show that the proposed methods perform well in small samples. The parameter estimators are virtually unbiased. The standard error estimators are accurate, and the confidence intervals have proper coverage probabilities.

## 6. Examples

To illustrate the methods of sections 2 and 3 and to compare with the existing methods of Lin & Wei (1992) and Lin *et al.* (1998), we consider the well-known bladder cancer data reported by Wei *et al.* (1989). These data were obtained from a randomized clinical trial assessing the potential benefit of thiotepa in reducing recurrences of bladder tumours. There are 38 patients in the thiotepa group with a total of 45 observed recurrences and 48 placebo patients with a total of 87 observed recurrences. To compare with the results of Lin & Wei (1992), we consider the first three recurrences of each patient. For $i = 1, \ldots, 86$ and $k = 1, 2, 3$, let $T_{ki}$ be the time from the initiation of treatment to the $k$th tumour recurrence on the $i$th patient, let $X_{1ki}$ indicated by the values 1 versus 0 whether the $i$th patient received thiotepa or placebo, and let $X_{2ki}$ be the number of initial tumours for the $i$th patient. We regress $\log_{10} T_{ki}$ on $X_{1ki}$ and $X_{2ki}$. Recurrence times of 0 are replaced with 0.5. In this section, the log-rank estimates at convergence are reported, and the resampling was performed in the same manner as in section 5 except that 10,000 samples are used. The results of our analysis are presented in Table 3. The log-rank estimates for individual recurrences are similar to those of Lin & Wei (1992). The optimally combined log-rank estimate is about 10% smaller than the estimate of Lin & Wei (1992) based on minimum-dispersion statistics. More importantly, our confidence intervals are much narrower than Lin & Wei's. In fact, our two-sided $p$-value for testing no overall treatment effect is approximately 0.039 whereas that of Lin & Wei (1992) is approximately 0.05. These differences reflect the fact that the Lin–Wei estimator is not based on the optimal linear combination.

Following Lin *et al.* (1998), we regard the tumour recurrences for each patient as a single counting process and fit model (4) with three covariates: (i) treatment indicator; (ii) number of initial tumours; and (iii) the diameter of the largest initial tumour; the treatment indicator takes the value 1 for placebo and 0 for thiotepa. Table 4 displays the results of our analysis, which are similar to those of Lin *et al.* (1998). Incidentally, Lin *et al.* (1998) used *ad hoc* iterative (one-dimensional) bisection search to solve the estimating functions along with a different resampling technique.

Table 3. *Estimation of treatment effects on the first three tumor recurrences of bladder cancer patients*

| Weight function | Tumour recurrences | Parameter estimate | Estimated standard error | 95% Wald confidence interval |
|---|---|---|---|---|
| Gehan | First | 0.289 | 0.205 | −0.114 to 0.691 |
| | Second | 0.302 | 0.126 | 0.055 to 0.549 |
| | Third | 0.246 | 0.126 | −0.001 to 0.492 |
| | First three | 0.272 | 0.120 | 0.037 to 0.506 |
| Log-rank | First | 0.392 | 0.213 | −0.026 to 0.809 |
| | Second | 0.295 | 0.151 | −0.002 to 0.592 |
| | Third | 0.248 | 0.127 | −0.001 to 0.497 |
| | First three | 0.260 | 0.126 | 0.013 to 0.508 |

The optimal linear estimator is used to estimate the overall treatment effect on the first three recurrences.

Table 4. *Regression analysis on the mean frequency of tumour recurrences in bladder cancer patients*

| Weight function | Covariate | Parameter estimate | Estimated standard error | 95% confidence intervals | |
|---|---|---|---|---|---|
| | | | | Wald | Percentile |
| Gehan | Treatment | 0.658 | 0.300 | 0.070 to 1.246 | 0.108 to 1.264 |
| | Initial number | 0.218 | 0.093 | 0.035 to 0.400 | 0.079 to 0.434 |
| | Initial size | −0.023 | 0.098 | −0.215 to 0.170 | −0.195 to 0.195 |
| Logrank | Treatment | 0.542 | 0.292 | −0.029 to 1.114 | 0.052 to 1.167 |
| | Initial number | 0.204 | 0.077 | 0.054 to 0.354 | 0.097 to 0.391 |
| | Initial size | −0.038 | 0.085 | −0.204 to 0.127 | −0.198 to 0.141 |

We change $\boldsymbol{\beta}$ to $-\boldsymbol{\beta}$ so as to be consistent with the parameterization of Lin *et al.* (1998).

For a real example of clustered data, we consider the litter-matched tumorigenesis data originally reported by Mantel *et al.* (1977) and reproduced in Table 1 of Lee *et al.* (1993). There are 50 female litters in the study, each having three rats. For $i=1,\ldots,50$ and $k=1,2,3$, let $T_{ik}$ be the time of tumour appearance for the $k$th rat in the $i$th litter, and let $X_{ik}$ indicate, by the values 0 versus 1, whether the $k$th rat in the $i$th litter was drug-treated or not. We regress $\log T_{ik}$ on $X_{ik}$. The Gehan estimate is 0.156 with an estimated standard error of 0.093, and the corresponding Wald 95% confidence interval is $(-0.026, 0.338)$. The log-rank estimate is 0.161 with an estimated standard error of 0.090, and the corresponding Wald 95% confidence interval is $(-0.016, 0.338)$. The log-rank results differ slightly from those of Lee *et al.* (1993).

## 7. Discussion

Although Cox-type regression methods for multivariate failure time data have been studied extensively, it is desirable to explore the accelerated failure time regression approach for several reasons. First, accelerated failure time models may fit the data better than proportional hazards models. Secondly, the accelerated failure time model formulates a natural and direct regression relationship, whereas the relative risk modelled by the Cox regression has no physical interpretation when the censored response variable is not failure time. Thirdly, the regression parameters under multivariate accelerated failure time models have both the population-averaged and subject-specific interpretations. This is not true of proportional hazards models.

The proposed resampling approach differs from that of Jin *et al.* (2003) and entails considerable technical challenges. The fact that this approach correctly adjusts for the intraclass dependence is remarkable. In all the existing methods for multivariate failure time data, either under proportional hazards models or accelerated failure time models, each estimating function is approximated by a sum of independent and identically distributed (i.i.d.) terms and the empirical variances and covariances of these sums are calculated. These calculations led to complicated variance–covariance expressions, which may perform poorly in small samples. The proposed resampling procedure does not involve complicated i.i.d. approximations in the variance–covariance estimation.

We have focused on the estimation of the regression parameters. A related problem is the estimation of the failure time distributions. The cumulative hazard functions for multiple events and clustered data as well as the mean frequency functions for recurrent events can be estimated consistently by the Aalen–Breslow type estimators (see Lin *et al.*, 1998, p. 608). Upon normalizations, these estimators converge weakly to zero-mean Gaussian processes. We can approximate the limiting distributions by extending the resampling technique developed in this paper, and construct appropriate confidence intervals and confidence bands.

Residuals similar to those of proportional hazards models (Kalbfleisch & Prentice, 2002, pp. 210–212) can be used to check accelerated failure time models. It is also possible to develop formal goodness-of-fit methods based on the comparison of the rank-type estimators with different weight functions (Wei *et al.*, 1990) or on the cumulative sums of residuals (Lin *et al.*, 1993). The resampling technique presented in this paper will be useful in evaluating the distributions of the test statistics.

## Acknowledgements

This research was supported by the New York City Council Speaker's Fund for Public Health Research, the National Institutes of Health and the National Science Foundation.

## References

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* **52**, 203–223.

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

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The statistical analysis of failure time data*, 2nd edn. John Wiley, Hoboken, NJ.

Lai, T. L. & Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531–556.

Lee, E. W., Wei, L. J. & Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Amer. Statist. Assoc.* **88**, 557–565.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Med.* **13**, 2233–2247.

Lin, J. S. & Wei, L. J. (1992). Linear regression analysis for multivariate failure time observations. *J. Amer. Statist. Assoc.* **87**, 1091–1097.

Lin, D. Y., Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

Lin, D. Y., Wei, L. J. & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711–730.

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

Pollard, D. (1990). *Empirical processes: theory and application*. IMS, Hayward, CA.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley, New York, NY.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–372.

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

Wei, L. J., Ying, Z. & Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.

*Received June 2004, in final form September 2005*

D. Y. Lin, Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill, NC 27599-7420, USA.
E-mail: lin@bios.unc.edu

## Appendix

*Proofs of asymptotic results*

The proofs in this appendix are more technical and more rigorous than those of Jin *et al.* (2003). We omit the kind of derivation given in the appendix of Jin *et al.* (2003) and focus on new technical issues. We first state and prove a lemma that is used repeatedly in our proofs.

**Lemma 1**

*Let $\mathbf{H}_n(t)$ and $W_n(t)$ be two sequences of bounded processes. Suppose that $\mathbf{H}_n(t)$ is component-wise monotone and converges in probability to $\mathbf{H}(t)$ uniformly in t and $W_n(t)$ converges weakly to a zero-mean process with continuous sample paths. Then for any continuously differentiable function g,*

$$\int_{-\infty}^{\infty}[g\{\mathbf{H}_n(t)\}-g\{\mathbf{H}(t)\}]\,\mathrm{d}W_n(t)=o_p(1).$$

*Proof of lemma 1*

By the strong embedding arguments as used in Lin *et al.* (2000, p. 726), $\mathbf{H}_n$ and $W_n$ can be assumed to converge to their respective limits almost surely. One can then apply lemma 1 of Lin *et al.* (2000) repeatedly and component-wise to obtain the desired approximation.

*Proof of theorem 1*

The classical strong law of large numbers for $U$ statistics (Serfling, 1980, section 5.4) implies that, under condition 1, $L_{k,G}$ converges almost surely for each $k$. Note that $L_{k,G}$ is a convex function, so that the convergence is uniform in any compact region. By condition 2, the limiting function is strictly convex at the true parameter value $\boldsymbol{\beta}_k$. Therefore, $\widehat{\boldsymbol{\beta}}_{k,G}\xrightarrow{a.s.}\boldsymbol{\beta}_k$.

Under conditions 1–3, we can apply the arguments of Ying (1993) to obtain

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{k,G}-\boldsymbol{\beta}_k)=-\mathbf{A}_{k,G}^{-1}n^{-1/2}\mathbf{U}_{k,G}(\boldsymbol{\beta}_k)+o(1+n^{1/2}||\widehat{\boldsymbol{\beta}}_{k,G}-\boldsymbol{\beta}_k||),\quad a.s. \tag{7}$$

Recall that

$$\mathbf{U}_{k,G}(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}S_k^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t).$$

The simple equality

$$\sum_{i=1}^{n}I\{e_{ki}(\boldsymbol{\beta})\geq t\}\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}=\mathbf{0}$$

implies that

$$\mathbf{U}_{k,G}(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}S_k^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}M_{ki}(\boldsymbol{\beta};t), \tag{8}$$

where the $M_{ki}(\boldsymbol{\beta};t)$ are defined in (1). It is well known that $E\{M_{ki}(\boldsymbol{\beta}_k;t)\}=0$. By the uniform strong law of large numbers (Pollard, 1990, section 8), $\mathbf{S}_k^{(r)}(\boldsymbol{\beta};t)\xrightarrow{a.s.}\mathbf{s}_k^{(r)}(\boldsymbol{\beta};t)$ uniformly in $\boldsymbol{\beta}$ and $t$. It then follows from lemma 1 that

$$n^{-1/2}\mathbf{U}_{k,G}(\boldsymbol{\beta}_k)=n^{-1/2}\sum_{i=1}^{n}\mathbf{u}_{ki,\,G}+o_p(1), \tag{9}$$

where

$$\mathbf{u}_{ki,\,G} = \int_{-\infty}^{\infty} s_k^{(0)}(\boldsymbol{\beta}_k;t)\{\mathbf{X}_{ki} - \bar{\mathbf{x}}_k(t)\}\,\mathrm{d}M_{ki}(\boldsymbol{\beta}_k;t).$$

In view of (7) and (9), the convergence of $n^{1/2}(\widehat{\mathbf{B}}_G - \mathbf{B})$ stated in theorem 1 follows from the multivariate central limit theorem.

Because of the way the random perturbation is introduced, the loss function $L^*_{k,G}$ retains the convexity of $L_{k,G}$. Thus, the above arguments for the consistency of $\widehat{\boldsymbol{\beta}}_{k,G}$ can be used to show that $\widehat{\boldsymbol{\beta}}^*_{k,G} \xrightarrow{a.s.} \boldsymbol{\beta}_k$.

Through algebraic manipulations, we can express (3) as

$$\mathbf{U}^*_{k,G}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \tilde{S}_k^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)Z_i, \tag{10}$$

where

$$\tilde{\mathbf{S}}_k^{(r)}(\boldsymbol{\beta};t) = n^{-1}\sum_{j=1}^{n} I\{e_{kj}(\boldsymbol{\beta}) \geq t\}\mathbf{X}_{kj}^r Z_j \ (r=0,1),$$

and

$$\overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t) = \tilde{\mathbf{S}}_k^{(1)}(\boldsymbol{\beta};t)/\tilde{S}_k^{(0)}(\boldsymbol{\beta};t).$$

This is a functional of weighted empirical processes, just like $\mathbf{U}_{k,G}$, but with the extra weights $Z_i$. Thus, the arguments for establishing the asymptotic linearity of $\mathbf{U}_{k,G}$ are applicable to $\mathbf{U}^*_{k,G}$ under conditions 1–3. In particular, we can expand $\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}^*_{k,G})$ at $\widehat{\boldsymbol{\beta}}_{k,G}$ to obtain

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^*_{k,G} - \widehat{\boldsymbol{\beta}}_{k,G}) = -\mathbf{A}_{k,G}^{-1} n^{-1/2}\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G}) + o(1 + n^{1/2}\|\widehat{\boldsymbol{\beta}}^*_{k,G} - \widehat{\boldsymbol{\beta}}_{k,G}\|), \quad a.s. \tag{11}$$

Note that (7) and (11) have the same slope matrix $\mathbf{A}_{k,G}$. This is because $n^{-1}\mathbf{U}_{k,G}(\boldsymbol{\beta})$ and $n^{-1}\mathbf{U}^*_{k,G}(\boldsymbol{\beta})$ converge to the same limiting function as $Z_i$ are independent of the data with mean 1.

As

$$\sum_{i=1}^{n} I\{e_{ki}(\boldsymbol{\beta}) \geq t\} Z_i\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\} = \mathbf{0},$$

we can rewrite (10) as

$$\mathbf{U}^*_{k,G}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \tilde{S}_k^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\}\,\mathrm{d}M_{ki}(\boldsymbol{\beta};t)Z_i. \tag{12}$$

This result arises from the specific way in which the random weights $Z_i$ are introduced into $L^*_{k,G}(\boldsymbol{\beta})$, and does *not* hold under the weighting scheme of Jin *et al.* (2003). In fact, the latter would not lead to a correct approximation. We shall show that $\tilde{S}_k^{(0)}$ and $\overline{\mathbf{X}}^*_k$ in (12) can be replaced by $S_k^{(0)}$ and $\overline{\mathbf{X}}_k$. This part of the proof is much more delicate than its counterpart in Jin *et al.* (2003).

Simple algebraic manipulations of (12) yield the following decomposition:

$$\begin{aligned}\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G}) = &\sum_{i=1}^{n} \int_{-\infty}^{\infty} S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k,G};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k,G};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)Z_i\\ &+ \int_{-\infty}^{\infty} \{\tilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k,G};t) - S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k,G};t)\}\,\mathrm{d}\sum_{i=1}^{n}\mathbf{X}_{ki}M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)Z_i\\ &- \int_{-\infty}^{\infty} \{\tilde{\mathbf{S}}_k^{(1)}(\widehat{\boldsymbol{\beta}}_{k,G};t) - \mathbf{S}_k^{(1)}(\widehat{\boldsymbol{\beta}}_{k,G};t)\}\,\mathrm{d}\sum_{i=1}^{n} M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)Z_i.\end{aligned} \tag{13}$$

Let $\mathcal{F}_n$ be the $\sigma$-algebra generated by all potential data $(T_{ki}, C_{ki}, \mathbf{X}_{ki})$ $(k=1,\ldots,K; i=1,\ldots,n)$. For random vectors $\mathbf{W}_n$ involving the $Z_i$s, we use the notation $\mathbf{W}_n=\tilde{o}(d_n)$ to denote the fact that $\Pr(\|d_n^{-1}\mathbf{W}_n\|>\epsilon|\mathcal{F}_n)\xrightarrow{a.s.}0$ for every $\epsilon>0$. Conditional on $\mathcal{F}_n$, $n^{1/2}\{\tilde{\mathbf{S}}_k^{(r)}(\boldsymbol{\beta};\cdot)-\mathbf{S}_k^{(r)}(\boldsymbol{\beta};\cdot)\}$ $(r=0,1)$ converge weakly, and

$$n^{-1}\sum_{i=1}^{n} M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)Z_i\to 0$$

and

$$n^{-1}\sum_{i=1}^{n}\mathbf{X}_{ki}M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)Z_i\to\mathbf{0}$$

uniformly in $t$. It then follows from lemma 1, together with integration by parts that the second and third terms on the right-hand side of (13) are both of order $\tilde{o}(n^{1/2})$. Clearly,

$$\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G})=\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G})-\mathbf{U}_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G})+\tilde{o}(n^{1/2})$$

as $\widehat{\boldsymbol{\beta}}_{k,G}$ is a root of $\mathbf{U}_{k,G}(\boldsymbol{\beta})$. By subtracting (8) evaluated at $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}_{k,G}$ from the first term on the right-hand side of (13), we have

$$\mathbf{U}^*_{k,G}(\widehat{\boldsymbol{\beta}}_{k,G})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k,G};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k,G};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t)(Z_i-1)+\tilde{o}(n^{1/2}). \qquad (14)$$

Conditional on $\mathcal{F}_n$, the first term on the right-hand side of (14) is a sum of zero-mean random vectors. Thus, the multivariate central limit theorem implies that the conditional distribution of the random vector $n^{-1/2}\{\mathbf{U}^{*'}_{1,G}(\widehat{\boldsymbol{\beta}}_{1,G}),\ldots,\mathbf{U}^{*'}_{K,G}(\widehat{\boldsymbol{\beta}}_{K,G})\}'$ given $\mathcal{F}_n$ converges almost surely to a $pK$-variate normal random vector with mean zero and covariance matrix

$$\{\tilde{\mathbf{V}}_{kl,G}; k,l=1,\ldots,K\},$$

where

$$\tilde{\mathbf{V}}_{kl,G}=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{u}}_{ki,G}\tilde{\mathbf{u}}'_{li,G},$$

and

$$\tilde{\mathbf{u}}_{ki,G}=\int_{-\infty}^{\infty}S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k,G};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k,G};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k,G};t),\quad k=1,\ldots,K;\; i=1,\ldots,n.$$

As $\mathbf{S}_k^{(r)}(\boldsymbol{\beta};t)\xrightarrow{a.s.}\mathbf{s}_k^{(r)}(\boldsymbol{\beta};t)$ $(r=0,1)$ and $\widehat{\boldsymbol{\beta}}_{k,G}\xrightarrow{a.s.}\boldsymbol{\beta}_k$, we have $\tilde{\mathbf{V}}_{kl,G}\xrightarrow{a.s.}\mathbf{V}_{kl,G}$. It then follows from (11) that the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^{*'}_{1,G}-\widehat{\boldsymbol{\beta}}'_1,\ldots,\widehat{\boldsymbol{\beta}}^{*'}_{K,G}-\widehat{\boldsymbol{\beta}}'_K)'$ given $\mathcal{F}_n$ converges almost surely to a zero-mean normal distribution with covariance matrix $\{\mathbf{A}_{k,G}^{-1}\mathbf{V}_{kl,G}\mathbf{A}_{l,G}^{-1}; k,l=1,\ldots,K\}$, which is the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}'_{1,G}-\boldsymbol{\beta}'_1,\ldots,\widehat{\boldsymbol{\beta}}'_{K,G}-\boldsymbol{\beta}'_K)'$.

*Proof of theorem 2*

The convex analysis arguments for establishing the consistency of $\widehat{\boldsymbol{\beta}}_{k,G}$ and $\widehat{\boldsymbol{\beta}}^*_{k,G}$ in the proof of theorem 1 can be used repeatedly to show that both $\widehat{\boldsymbol{\beta}}_{k(m)}$ and $\widehat{\boldsymbol{\beta}}^*_{k(m)}$ are strongly consistent. To derive the asymptotic distributions, we note that $\widehat{\boldsymbol{\beta}}_{k(m)}$ and $\widehat{\boldsymbol{\beta}}^*_{k(m)}$ are the roots of

$$\tilde{\mathbf{U}}_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{k(m-1)})\equiv\sum_{i=1}^{n}\int_{-\infty}^{\infty}\psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t+(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{k(m-1)})'\mathbf{X}_{ki})$$
$$\times S_k^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_{ki}-\overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t),$$

$$\tilde{\mathbf{U}}^*_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}^*_{k(m-1)}) \equiv \sum_{i=1}^n \int_{-\infty}^{\infty} \psi_k(\widehat{\boldsymbol{\beta}}^*_{k(m-1)}; t + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*_{k(m-1)})'\mathbf{X}_{ki}) \times \tilde{S}^{(0)}_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)Z_i$$

respectively. Under condition 5,

$$\begin{aligned}\tilde{\mathbf{U}}_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{k(m-1)}) &= \sum_{i=1}^n \int_{-\infty}^{\infty} \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t) S^{(0)}_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)\\ &\quad + \sum_{i=1}^n \int_{-\infty}^{\infty} \dot{\psi}_{0k}(t) S^{(0)}_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)\\ &\quad \times (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{k(m-1)})'\mathbf{X}_{ki} + o(n^{1/2} + n\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{k(m-1)}\|)\end{aligned} \tag{15}$$

$$\begin{aligned}\tilde{\mathbf{U}}^*_{k,\phi_k}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}^*_{k(m-1)}) &= \sum_{i=1}^n \int_{-\infty}^{\infty} \psi_k(\widehat{\boldsymbol{\beta}}^*_{k(m-1)};t) \tilde{S}^{(0)}_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)Z_i\\ &\quad + \sum_{i=1}^n \int_{-\infty}^{\infty} \dot{\psi}_{0k}(t) \tilde{S}^{(0)}_k(\boldsymbol{\beta};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\boldsymbol{\beta};t)\}\,\mathrm{d}N_{ki}(\boldsymbol{\beta};t)\\ &\quad \times Z_i(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*_{k(m-1)})'\mathbf{X}_{ki} + o(n^{1/2} + n\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*_{k(m-1)}\|).\end{aligned} \tag{16}$$

Given (7) and (15), we can extend the arguments for establishing (11) of Jin *et al.* (2003) to show that the following result holds under conditions 1–5,

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\beta}}_{k(m)} - \boldsymbol{\beta}_k) &= -n^{-1/2}[\mathbf{I} - \{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m]\mathbf{A}^{-1}_{k,\phi_k}\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta}_k)\\ &\quad - n^{-1/2}\{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m \mathbf{A}^{-1}_{k,G}\mathbf{U}_{k,G}(\boldsymbol{\beta}_k)\\ &\quad + o\left(1 + n^{1/2}\sum_{j=0}^m ||\widehat{\boldsymbol{\beta}}_{k(j)} - \boldsymbol{\beta}_k||\right).\end{aligned}$$

Note that condition 4 is necessary for the above equation to be meaningful. By the arguments for establishing (9) in the proof of theorem 1, we have

$$n^{-1/2}\mathbf{U}_{k,\phi_k}(\boldsymbol{\beta}_k) = n^{-1/2}\sum_{i=1}^n \mathbf{u}_{ki,\phi_k} + o_p(1),$$

where the $\mathbf{u}_{ki,\phi_k}$ are defined in (2). Thus,

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\beta}}_{k(m)} - \boldsymbol{\beta}_k) &= -n^{-1/2}\sum_{i=1}^n \Big([\mathbf{I} - \{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m]\mathbf{A}^{-1}_{k,\phi_k}\mathbf{u}_{ki,\phi_k}\\ &\quad + \{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m \mathbf{A}^{-1}_{k,G}\mathbf{u}_{ki,G}\Big)\\ &\quad + o\left(1 + n^{1/2}\sum_{j=0}^m ||\widehat{\boldsymbol{\beta}}_{k(j)} - \boldsymbol{\beta}_k||\right).\end{aligned} \tag{17}$$

In analogy to (12) of Jin *et al.* (2003), the following result follows from (16)

$$\begin{aligned}\tilde{\mathbf{U}}^*_{k,\phi_k}(\widehat{\boldsymbol{\beta}}^*_{k(m)};\widehat{\boldsymbol{\beta}}^*_{k(m-1)}) &= \sum_{i=1}^n \int_{-\infty}^{\infty} \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)}; t + (\widehat{\boldsymbol{\beta}}_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m-1)})'\mathbf{X}_{ki})\\ &\quad \times \tilde{S}^{(0)}_k(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}^*_k(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}N_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t)Z_i\\ &\quad + n(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})(\widehat{\boldsymbol{\beta}}^*_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m)})\\ &\quad - n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}^*_{k(m-1)} - \widehat{\boldsymbol{\beta}}_{k(m-1)}) + \mathrm{d}^*_k,\end{aligned} \tag{18}$$

where

$$d_k^* = \tilde{o}\left(n^{1/2} + n\sum_{j=0}^{m}\{||\widehat{\boldsymbol{\beta}}_{k(j)} - \boldsymbol{\beta}_k|| + ||\widehat{\boldsymbol{\beta}}^*_{k(j)} - \boldsymbol{\beta}_k||\}\right).$$

Under condition 5, up to an asymptotically negligible term, the first term on the right-hand side of (18) can be written as

$$\sum \int \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t)\tilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k^*(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}N_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t)Z_i + n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m-1)}),$$

or

$$\sum \int \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t)\tilde{S}_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k^*(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t)Z_i + n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m-1)}),$$

which, up to order $\tilde{o}(n^{1/2})$, is equivalent to

$$\sum \int \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t)S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t)Z_i + n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m-1)}).$$

The equivalence between the last two expressions follows from a decomposition similar to (13). On the other hand, $\tilde{\mathbf{U}}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)};\widehat{\boldsymbol{\beta}}_{k(m-1)})$ can be expressed as

$$\sum \int \psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t)S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t) + n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m-1)})$$

plus an asymptotically negligible term. Thus, the subtraction of $\tilde{\mathbf{U}}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}_{k(m)};\widehat{\boldsymbol{\beta}}_{k(m-1)})$ from the right-hand side of (18) yields

$$\tilde{\mathbf{U}}^*_{k,\phi_k}(\widehat{\boldsymbol{\beta}}^*_{k(m)};\widehat{\boldsymbol{\beta}}^*_{k(m-1)}) = \sum_{i=1}^{n}\tilde{\mathbf{u}}_{ki,\phi_k}(Z_i - 1) + n(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})(\widehat{\boldsymbol{\beta}}^*_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m)}) - n\mathbf{D}_{k,\phi_k}(\widehat{\boldsymbol{\beta}}^*_{k(m-1)} - \widehat{\boldsymbol{\beta}}_{k(m-1)}) + \mathrm{d}_k^*,$$

where

$$\tilde{\mathbf{u}}_{ki,\phi_k} = \int_{-\infty}^{\infty}\psi_k(\widehat{\boldsymbol{\beta}}_{k(m-1)};t)S_k^{(0)}(\widehat{\boldsymbol{\beta}}_{k(m)};t)\{\mathbf{X}_{ki} - \overline{\mathbf{X}}_k(\widehat{\boldsymbol{\beta}}_{k(m)};t)\}\,\mathrm{d}M_{ki}(\widehat{\boldsymbol{\beta}}_{k(m)};t).$$

Therefore,

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^*_{k(m)} - \widehat{\boldsymbol{\beta}}_{k(m)}) = -n^{-1/2}\sum_{i=1}^{n}\Big([\mathbf{I} - \{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m]\mathbf{A}^{-1}_{k,\phi_k}\tilde{\mathbf{u}}_{ki,\phi_k} + \{(\mathbf{A}_{k,\phi_k} + \mathbf{D}_{k,\phi_k})^{-1}\mathbf{D}_{k,\phi_k}\}^m\mathbf{A}^{-1}_{k,G}\tilde{\mathbf{u}}_{ki,\,G}\Big)(Z_i - 1) + n^{-1/2}\,\mathrm{d}_k^*. \tag{19}$$

By comparing (17) and (19), we conclude that the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^{*'}_{1(m)} - \widehat{\boldsymbol{\beta}}'_{1(m)}, \ldots, \widehat{\boldsymbol{\beta}}^{*'}_{K(m)} - \widehat{\boldsymbol{\beta}}'_{K(m)})'$ given $\mathcal{F}_n$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}'_{1(m)} - \boldsymbol{\beta}'_1, \ldots, \widehat{\boldsymbol{\beta}}'_{K(m)} - \boldsymbol{\beta}'_K)'$.

*Proof of theorem 3*

As in the proof of theorem 1, the convexity of the loss functions, $L_G$ and $L^*_G$, together with the non-singularity of the second derivative of their common limit under condition 8, implies that both $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}^*_G$ are strongly consistent.

We can express (5) as

$$\mathbf{U}_{\phi}(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{0}^{\infty}\phi(\boldsymbol{\beta};t)\{\mathbf{X}_i-\overline{\mathbf{X}}(\boldsymbol{\beta};t)\}\,\mathrm{d}M_i(\boldsymbol{\beta};t).$$

Under model (4), $E\{M_i(\boldsymbol{\beta}_0;t)\}=0$ $(i=1,\ldots,n)$. It follows from the functional central limit theorem (Pollard, 1990, p. 53) that $n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_iM_i(\boldsymbol{\beta}_0;\cdot)$ and $n^{-1/2}\sum_{i=1}^{n}M_i(\boldsymbol{\beta}_0;\cdot)$ converge to zero-mean Gaussian processes. By the uniform strong law of large numbers, $\mathbf{S}^{(r)}(\boldsymbol{\beta};t)\xrightarrow{a.s.}s^{(r)}(\boldsymbol{\beta};t)$ $(r=0,1)$ uniformly in $\boldsymbol{\beta}$ and $t$. It then follows from lemma 1 that $n^{-1/2}\mathbf{U}_{\phi}(\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}\mathbf{u}_{i,\phi}+o_p(1)$. In view of this equation, the multivariate central limit theorem implies that $n^{-1/2}\mathbf{U}_G(\boldsymbol{\beta}_0)$ converges weakly to a zero-mean normal random vector with covariance matrix $\mathbf{V}_G$.

It can be shown through algebraic manipulations that the derivative of $L_G^*(\boldsymbol{\beta})$ takes the form

$$\mathbf{U}_G^*(\boldsymbol{\beta})=\sum_{i=1}^{n}\int_{0}^{\infty}\tilde{S}^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_i-\overline{\mathbf{X}}^*(\boldsymbol{\beta};t)\}\,\mathrm{d}M_i(\boldsymbol{\beta};t)Z_i,$$

where $\overline{\mathbf{X}}^*(\boldsymbol{\beta};t)=\tilde{\mathbf{S}}^{(1)}(\boldsymbol{\beta};t)/\tilde{S}^{(0)}(\boldsymbol{\beta};t)$, and $\tilde{\mathbf{S}}^{(r)}(\boldsymbol{\beta};t)=n^{-1}\sum_{j=1}^{n}I(C_j\,\mathrm{e}^{-\boldsymbol{\beta}'\mathbf{X}_j}\geq t)\mathbf{X}_j^{r}Z_j$ $(r=0,1)$. Let $\mathcal{F}_n$ denote the $\sigma$-algebra generated by $(C_i,T_{ki},\mathbf{X}_i)$ $(T_{ki}\leq C_i;i=1,\ldots,n)$. By the arguments leading to (14),

$$\mathbf{U}_G^*(\widehat{\boldsymbol{\beta}}_G)=\sum_{i=1}^{n}\int_{0}^{\infty}\mathbf{S}^{(0)}(\widehat{\boldsymbol{\beta}}_G;t)\{\mathbf{X}_i-\overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_G;t)\}\,\mathrm{d}M_i(\widehat{\boldsymbol{\beta}}_G;t)(Z_i-1)+\tilde{o}(n^{1/2}).\qquad(20)$$

Thus, the multivariate central limit theorem implies that the conditional distribution of $n^{-1/2}\mathbf{U}_G^*(\widehat{\boldsymbol{\beta}}_G)$ converges almost surely to a zero-mean normal random vector with covariance matrix $\mathbf{V}_G$.

As both $\mathbf{U}_G(\boldsymbol{\beta})$ and $\mathbf{U}_G^*(\boldsymbol{\beta})$ are functionals of empirical processes, we can establish under conditions 6–8 the asymptotic linearities for $\mathbf{U}_G(\boldsymbol{\beta})$ and $\mathbf{U}_G^*(\boldsymbol{\beta})$ similar to (7) and (11). As $E(Z_i)=1$ $(i=1,\ldots,n)$, the slope matrices in the two expansions are identical. It follows that the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G^*-\widehat{\boldsymbol{\beta}}_G)$ given $\mathcal{F}_n$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G-\boldsymbol{\beta}_0)$.

*Proof of theorem 4*

The strong consistency of $\widehat{\boldsymbol{\beta}}_{(m)}$ and $\widehat{\boldsymbol{\beta}}_{(m)}^*$ again follow from the convexity arguments. Note that $\widehat{\boldsymbol{\beta}}_{(m)}$ and $\widehat{\boldsymbol{\beta}}_{(m)}^*$ are the roots of

$$\tilde{\mathbf{U}}_{\phi}(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{(m-1)})\equiv\sum_{i=1}^{n}\int_{0}^{\infty}\psi(\widehat{\boldsymbol{\beta}}_{(m-1)};t\,\mathrm{e}^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{(m-1)})'\mathbf{X}_i})S^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_i-\overline{\mathbf{X}}(\boldsymbol{\beta};t)\}\,\mathrm{d}N_i(\boldsymbol{\beta};t),$$

$$\tilde{\mathbf{U}}_{\phi}^*(\boldsymbol{\beta};\widehat{\boldsymbol{\beta}}_{(m-1)}^*)\equiv\sum_{i=1}^{n}\int_{0}^{\infty}\psi(\widehat{\boldsymbol{\beta}}_{(m-1)}^*;t\,\mathrm{e}^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{(m-1)}^*)'\mathbf{X}_i})\tilde{S}^{(0)}(\boldsymbol{\beta};t)\{\mathbf{X}_i-\overline{\mathbf{X}}^*(\boldsymbol{\beta};t)\}\,\mathrm{d}N_i(\boldsymbol{\beta};t)Z_i$$

respectively. Under condition 10, we can take the expansion of $\psi$ with respect to its second argument at $t$. In doing so, the arguments for establishing (17) and (19), together with theorem 2 of Lin *et al.* (1998), can be used to show that

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}_0)=-n^{-1/2}\sum_{i=1}^{n}\Big([\mathbf{I}-\{(\mathbf{A}_{\phi}+\mathbf{D}_{\phi})^{-1}\mathbf{D}_{\phi}\}^m]\mathbf{A}_{\phi}^{-1}\mathbf{u}_{i,\phi}+\{(\mathbf{A}_{\phi}+\mathbf{D}_{\phi})^{-1}\mathbf{D}_{\phi}\}^m\mathbf{A}_G^{-1}\mathbf{u}_{i,G}\Big)+o\Big(1+n^{1/2}\sum_{j=0}^{m}||\widehat{\boldsymbol{\beta}}_{(j)}-\boldsymbol{\beta}_0||\Big),$$

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^*_{(m)} - \widehat{\boldsymbol{\beta}}_{(m)}) = -n^{-1/2}\sum_{i=1}^{n}\Big([\mathbf{I} - \{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m]\mathbf{A}_\phi^{-1}\tilde{\mathbf{u}}_{i,\phi} + \{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m \mathbf{A}_G^{-1}\tilde{\mathbf{u}}_{i,G}\Big)(Z_i - 1) + \tilde{o}\Big(1 + n^{1/2}\sum_{j=0}^{m}\{||\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_0|| + ||\widehat{\boldsymbol{\beta}}^*_{(j)} - \boldsymbol{\beta}_0||\}\Big),$$

where $\tilde{\mathbf{u}}_{i,\phi} = \int_0^\infty \psi(\widehat{\boldsymbol{\beta}}_{(m-1)}; t)S^{(0)}(\widehat{\boldsymbol{\beta}}_{(m)}; t)\{\mathbf{X}_i - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_{(m)}; t)\}\,\mathrm{d}M_i(\widehat{\boldsymbol{\beta}}_{(m)}; t)$. Thus, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^*_{(m)} - \widehat{\boldsymbol{\beta}}_{(m)})$ given $\mathcal{F}_n$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0)$.

*Proof of theorem 5*

Under conditions 11–13, the strong consistency of $\widehat{\boldsymbol{\beta}}_G$ and $\widehat{\boldsymbol{\beta}}^*_G$ follows from the same arguments as in the proofs of theorem 1. We may express (6) as

$$\mathbf{U}_\phi(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{-\infty}^{\infty}\phi(\boldsymbol{\beta}; t)\{\mathbf{X}_{ik} - \overline{\mathbf{X}}(\boldsymbol{\beta}; t)\}\,\mathrm{d}M_{ik}(\boldsymbol{\beta}; t).$$

It is simple to show that the $M_{ik}(\boldsymbol{\beta}_0; \cdot)$ are zero-mean processes. The uniform law of large numbers, together with lemma 1, entails that $n^{-1/2}\mathbf{U}_\phi(\boldsymbol{\beta}_0) = \sum_{i=1}^{n}\sum_{k=1}^{K_i}\mathbf{u}_{ik,\phi} + o_p(1)$. Consequently, $n^{-1/2}\mathbf{U}_G(\boldsymbol{\beta}_0)$ is asymptotically zero-mean normal with covariance matrix $\mathbf{V}_G$.

The derivative of $L^*_G(\boldsymbol{\beta})$ can be written as

$$\mathbf{U}^*_G(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{-\infty}^{\infty}\tilde{S}^{(0)}(\boldsymbol{\beta}; t)\{\mathbf{X}_{ik} - \overline{\mathbf{X}}^*(\boldsymbol{\beta}; t)\}\,\mathrm{d}M_{ik}(\boldsymbol{\beta}; t)Z_i,$$

where $\tilde{\mathbf{S}}^{(r)}(\boldsymbol{\beta}; t) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i} I\{e_{ik}(\boldsymbol{\beta}) \geq t\}\mathbf{X}_{ik}^r Z_i$ $(r = 0, 1)$, and $\overline{\mathbf{X}}^*(\boldsymbol{\beta}; t) = \tilde{\mathbf{S}}^{(1)}(\boldsymbol{\beta}; t)/\tilde{S}^{(0)}(\boldsymbol{\beta}; t)$. Let $\mathcal{F}_n$ denote the $\sigma$-algebra generated by $(T_{ik}, C_{ik}, \mathbf{X}_{ik})$ $(i = 1, \ldots, n; k = 1, \ldots, K_i)$. Analogous to (14) and (20),

$$\mathbf{U}^*_G(\widehat{\boldsymbol{\beta}}_G) = \sum_{i=1}^{n}\sum_{k=1}^{K_i} S^{(0)}(\widehat{\boldsymbol{\beta}}_G; t)\{\mathbf{X}_{ik} - \overline{\mathbf{X}}^*(\widehat{\boldsymbol{\beta}}_G; t)\}\,\mathrm{d}M_{ik}(\widehat{\boldsymbol{\beta}}_G; t)(Z_i - 1) + \tilde{o}(n^{1/2}).$$

It then follows from the multivariate central limit theorem that the conditional distribution of $n^{-1/2}\mathbf{U}^*_G(\widehat{\boldsymbol{\beta}}_G)$ given $\mathcal{F}_n$ converges almost surely to a zero-mean normal random vector with covariance matrix $\mathbf{V}_G$. Because $n^{-1}\mathbf{U}_G(\boldsymbol{\beta})$ and $n^{-1}\mathbf{U}^*_G(\boldsymbol{\beta})$ have the same asymptotic slope matrix, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^*_G - \widehat{\boldsymbol{\beta}}_G)$ given $\mathcal{F}_n$ converges almost surely to the asymptotic distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}_0)$.

*Proof of theorem 6*

The strong consistency of $\widehat{\boldsymbol{\beta}}_{(m)}$ and $\widehat{\boldsymbol{\beta}}^*_{(m)}$ follows from the proof of theorem 2. The following two equations, which are analogous to (17) and (19), hold under conditions 11–15,

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0) = -n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\Big([\mathbf{I} - \{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m]\mathbf{A}_\phi^{-1}\mathbf{u}_{ik,\phi} + \{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m\mathbf{A}_G^{-1}\mathbf{u}_{ik,G}\Big) + o\Big(1 + n^{1/2}\sum_{j=0}^{m}||\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_0||\Big),$$

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^{*}_{(m)} - \widehat{\boldsymbol{\beta}}_{(m)}) = -n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\Big([\mathbf{I} - \{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m]\mathbf{A}_\phi^{-1}\tilde{\mathbf{u}}_{ik,\phi}$$
$$+\{(\mathbf{A}_\phi + \mathbf{D}_\phi)^{-1}\mathbf{D}_\phi\}^m\mathbf{A}_G^{-1}\tilde{\mathbf{u}}_{ik,G}\Big)(Z_i - 1)$$
$$+\tilde{o}\left(1 + n^{1/2}\sum_{j=0}^{m}\{||\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_0|| + ||\widehat{\boldsymbol{\beta}}^{*}_{(j)} - \boldsymbol{\beta}_0||\}\right),$$

where $\tilde{\mathbf{u}}_{ik,\phi} = \int_{-\infty}^{\infty}\psi(\widehat{\boldsymbol{\beta}}_{(m-1)};t)S^{(0)}(\widehat{\boldsymbol{\beta}}_{(m)};t)\{\mathbf{X}_{ik} - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_{(m)};t)\}\,\mathrm{d}M_{ik}(\widehat{\boldsymbol{\beta}}_{(m)};t)$. It follows that the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}^{*}_{(m)} - \widehat{\boldsymbol{\beta}}_{(m)})$ given $\mathcal{F}_n$ converges almost surely to the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}_0)$.