Statistica Sinica **22** (2012), 317-336
doi:http://dx.doi.org/10.5705/ss.2009.261

JOINT ANALYSIS OF LONGITUDINAL DATA WITH DEPENDENT OBSERVATION TIMES

Xingqiu Zhao, Xingwei Tong and Liuquan Sun

The Hong Kong Polytechnic University, Beijing Normal University and Chinese Academy of Science

Abstract: This article discusses regression analysis of longitudinal data that often occur in medical follow-up studies and observational investigations. For the analysis of these data, most of the existing methods assume that observation times are independent of recurrent events completely, or given covariates, which may not be true in practice. We propose a joint modeling approach that uses a latent variable and a completely unspecified link function to characterize the correlations between the longitudinal response variable and the observation times. For inference about regression parameters, estimating equation approaches are developed without involving estimation for latent variables and the asymptotic properties of the resulting estimators are established. Methods for model checking are also presented. The performance of the proposed estimation procedures are evaluated through Monte Carlo simulations, and a data set from a bladder tumor study is analyzed as an illustrative example.

Key words and phrases: Estimating equation, informative observation times, joint modeling, latent variable, longitudinal data.

1. Introduction

The analysis of longitudinal data has recently attracted considerable attention. These data frequently occur in medical follow-up studies and observational investigations. For the analysis of longitudinal data, a number of methods have been developed, mostly under the assumption that the longitudinal response process and the observation process are independent completely, or given covariates. For example, Diggle, Liang, and Zeger (1994) presented an excellent summary about such commonly used methods as estimating equation and random-effect model approaches, and Lin and Ying (2001) and Welsh, Lin, and Carroll (2002) discussed general semiparametric regression analysis of longitudinal data when both observation times and the censoring times may depend on covariates.

A common situation where informative observation times occur is that these times are subject or response variable-dependent. For example, they may be hospitalization times of subjects in the study (Wang, Qin and Chiang (2001)). In a bladder cancer study, Sun and Wei (2000) and Zhang (2002) discussed a set of longitudinal data arising from a bladder cancer follow-up study conducted by the Veterans Administration Cooperative Urological Research Group; in this study, some patients had significantly more clinical visits than others and thus the occurrence of bladder tumors of a patient and the visit times may be related. Lipsitz et al. (2002) presented a set of longitudinal data from a study of children with acute lymphoblastic leukemia that involved correlated response and observation processes. The same could be true for other medical follow-up studies, but there is limited research on the analysis of longitudinal data when the longitudinal response process of interest may be correlated with the observation process given covariates, that is, the observation times may be informative. Sun et al. (2005) studied semiparametric models that allow observation times to be correlated with the longitudinal process; Sun, Sun, and Liu (2007) proposed a joint model for the longitudinal process and the observation process, where both processes may be correlated through a shared latent variable or frailty, and used the estimating equation approach to estimate the regression parameters; Liang, Lu, and Ying (2009) discussed a joint model through two random effects, where the relationship between the random effects is specified and a parametric distribution assumption for a random effect is required. The aim of this paper is to consider more general joint models for longitudinal data with dependent observation times, to develop an estimating equation approach for estimation of regression parameters, and to establish the asymptotic properties of the resulting estimates.

The remainder of this paper is organized as follows. Section 2 introduces notation and describes joint models for the longitudinal response process and the observation time process, where a latent variable and a completely unspecified link function are used to characterize the correlation between the two processes. In Section 3, an estimating equation approach is proposed for estimation of regression parameters and the asymptotic properties of the resulting estimates are established. In Section 4, we discuss the assessment of the models described in Section 2. Section 5 presents some results obtained from a simulation study of the finite-sample properties of the proposed inference procedure. In Section 6, we apply the proposed methods to a data set from a bladder tumor study. Some concluding remarks are made in Section 7.

2. Joint Modeling

Consider a longitudinal study, with $Y(t)$ as the longitudinal response variable of interest. Let X be the p-dimensional vector of covariates, C be the follow-up or censoring time, and $N(t)$ be the counting process for the number of the observation times before or at time t. The longitudinal process $Y(t)$ is observed only at time points where $N(t)$ jumps, for $t \leq C$. Let Z be an unobserved positive random variable that is independent of X. We assume that

$$E\{Y(t)|X, Z\} = \mu_0(t) + \beta' X + g(Z), \tag{2.1}$$

where $\mu_0(\cdot)$ is an unspecified continuous function, $g(\cdot)$ is a completely unspecified link function with $E\{g(Z)\} = 0$, and β is a vector of unknown regression parameters. For the observation process, we assume that $N(t)$ is a Poisson process with intensity function

$$\lambda(t|X, Z) = Z\lambda_0(t)\exp(\gamma' X), \tag{2.2}$$

where $\lambda_0(\cdot)$ is a completely unknown continuous baseline intensity function and γ is a vector of unknown regression parameters. Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. Let τ be the length of the study and take $\Lambda_0(\tau) = 1$ to avoid the identifiability issue. In addition, we assume that the censoring time C is independent of X and Z, and conditional on X and Z, $Y(\cdot)$ and $N(\cdot)$ are mutually independent.

Model (2.2) has been studied by several authors for the analysis of recurrent event data (e.g., Huang and Wang (2004); Wang, Qin and Chiang (2001)). Sun, Sun, and Liu (2007) discussed a special case of joint models (2.1) and (2.2) by specifying $g(Z) = Z - E(Z)$ or $g(Z) = -\{Z - E(Z)\}$. However, there does not seem to be research on the general joint models (2.1) and (2.2). From the proposed joint models it is obvious that, given covariates, the longitudinal response process $Y(t)$ and the observation process $N(t)$ can be correlated and that their relationship is partly determined by a link function of the latent variable Z, while the link function and the distributional form of Z are left unspecified. Our main goal here is to make inference about β. Toward this end, we develop an estimating approach in the next section.

3. Estimation Procedures

Suppose that a longitudinal study involves n subjects and

$$\{Y_i(t), X_i, Z_i, C_i, N_i(t), i = 1, \ldots, n\}$$

is a random sample of $\{Y(t), X, Z, C, N(t)\}$. Also, suppose that $N_i(t)$ is observed only at finite time points $T_{i1} < \cdots < T_{iK_i}$, where K_i denotes the total number of observation before or at the censoring time C_i for subject i, $i = 1, \ldots, n$.

For estimation of β, let

$$\bar{Y}_i = \int_0^\tau Y_i(t) I(C_i \geq t) dN_i(t).$$

Now since

$$\begin{aligned}E(\bar{Y}_i|X_i,Z_i) &= \int_0^\tau \{\mu_0(t)+\beta' X_i+g(Z_i)\}P(C_i\geq t)Z_i\exp(\gamma' X_i)d\Lambda_0(t)\\ &= \beta' X_i Z_i\exp(\gamma' X_i)E\{\Lambda_0(C_i)\}\\ &\quad +Z_i\exp(\gamma' X_i)\int_0^\tau \mu_0(t)P(C_i\geq t)d\Lambda_0(t)\\ &\quad +g(Z_i)Z_i\exp(\gamma' X_i)\int_0^\tau P(C_i\geq t)d\Lambda_0(t),\\ E(\bar{Y}_i|X_i) &= E(Z_i)E\{\Lambda_0(C_i)\}\exp(\gamma' X_i)\beta' X_i\\ &\quad +\exp(\gamma' X_i)\int_0^\tau [E(Z_i)\mu_0(t)+E\{g(Z_i)Z_i\}]P(C_i\geq t)d\Lambda_0(t).\end{aligned}$$

Let $X_{1i}=(1,X_i')'$, $\theta_1=\log E(Z)$, $\theta=(\theta_1,\gamma')'$, $\psi=E\{\Lambda_0(C)\}$, and

$$\alpha=[E\{\Lambda_0(C)\}]^{-1}\int_0^\tau \Big[\mu_0(t)+\frac{E\{g(Z)Z\}}{E(Z)}\Big]P(C\geq t)d\Lambda_0(t).$$

Then, we have

$$E\left\{\psi^{-1}\exp(-\theta' X_{1i})\bar{Y}_i-\alpha-\beta' X_i\right\}=0. \tag{3.1}$$

Motivated by (3.1), for given ψ and θ, we can consider the estimating function

$$U(\beta,\alpha;\psi,\theta)=\frac{1}{n}\sum_{i=1}^n W_i X_{1i}\left\{\psi^{-1}\exp(-\theta' X_{1i})\bar{Y}_i-\alpha-\beta' X_i\right\},$$

where W_i's are weights that could depend on the X_i's and C_i's. Let $\tilde{\beta}$ and $\tilde{\alpha}$ denote the solution to $U(\alpha,\beta;\psi,\theta)=0$. Then,

$$\begin{pmatrix}\tilde{\alpha}\\ \tilde{\beta}\end{pmatrix}=\left\{\sum_{i=1}^n W_i X_{1i}^{\otimes 2}\right\}^{-1}\sum_{i=1}^n W_i X_{1i}\psi^{-1}\exp(-\theta' X_{1i})\bar{Y}_i,$$

where $a^{\otimes 2}=aa'$ for vector a.

Of course ψ and θ are unknown and we cannot directly use the estimating function $U(\beta,\alpha;\psi,\theta)$. For this, we first consider inference about model (2.2). Let $\{s_\ell,\ell=1,\ldots,m\}$ denote the ordered and distinct values of all observation times $\{T_{ij},j=1,\ldots,K_i,i=1,\ldots,n\}$, $q_\ell=\sum_{i=1}^n dN_i(s_\ell)$ be the number of observations of s_ℓ, and $N_\ell=\sum_{i=1}^n I(s_\ell\leq C_i)N_i(s_\ell)$ be the total number of observations with observation times and censoring time satisfying $T_{ij}\leq s_\ell\leq C_i$. Then we can derive the conditional likelihood function of the observed data on the N_i's conditional on $\{K_i,C_i,X_i,Z_i\}$, and the nonparametric maximum likelihood estimator $\hat{\Lambda}_0(t)$ of $\Lambda_0(t)$ given by

$$\hat{\Lambda}_0(t)=\prod_{s_\ell>t}(1-\frac{q_\ell}{N_\ell})$$

(Wang, Qin and Chiang (2001)), where the product is taken to be 1 if there is no s_ℓ with $s_\ell > t$. Thus, a natural estimator of ψ is given by $\hat{\psi} = n^{-1}\sum_{i=1}^n \hat{\Lambda}_0(C_i)$. It is easy to show that $\hat{\psi}$ is consistent.

For estimation of θ, Wang, Qin and Chiang (2001) proposed using the estimating equation

$$\frac{1}{n}\sum_{i=1}^n \eta_i X_{1i}\left\{K_i\hat{\Lambda}_0^{-1}(C_i) - \exp(\theta' X_{1i})\right\} = 0, \tag{3.2}$$

where η_i is a weight function that could depend on $(X_i, \theta, \hat{\Lambda}_0)$. The solution to (3.2) is denoted by $\hat{\theta}$.

We propose estimating α and β by using estimating function $U(\alpha, \beta; \hat{\psi}, \hat{\theta})$. Let $\hat{\alpha}$ and $\hat{\beta}$ denote the solution to $U(\alpha, \beta; \hat{\psi}, \hat{\theta}) = 0$. Then,

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \left\{\sum_{i=1}^n W_i X_{1i}^{\otimes 2}\right\}^{-1}\sum_{i=1}^n W_i X_{1i}\hat{\psi}^{-1}\exp(-\hat{\theta}' X_{1i})\bar{Y}_i.$$

It is easy to show from the Law of Large Numbers and the consistency of $\hat{\psi}$, $\hat{\Lambda}_0$, and $\hat{\theta}$ that the estimators $\hat{\alpha}$ and $\hat{\beta}$ are consistent.

To establish the asymptotic normality of $\hat{\alpha}$ and $\hat{\beta}$, let

$$\begin{aligned}
H_n(t) &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{K_i} I(T_{ij} \le t),\\
R_n(t) &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{K_i} I(T_{ij} \le t \le C_i),\\
b_{in}(t) &= \sum_{j=1}^{K_i}\left\{\int_t^\tau \frac{I(T_{ij} \le u \le C_i)dH_n(u)}{R_n^2(u)} - \frac{I(t < T_{ij} \le \tau)}{R_n(T_{ij})}\right\},\\
\hat{e}_{in} &= -\frac{1}{n}\sum_{j=1}^n \eta_j X_{1j}K_j b_{in}(C_j)\{\hat{\Lambda}_0(C_j)\}^{-1} + \eta_i X_{1i}[K_i\{\hat{\Lambda}_0(C_i)\}^{-1} - \exp(\hat{\theta}' X_{1i})],\\
\hat{f}_{in} &= \left\{n^{-1}\sum_{j=1}^n \eta_j X_{1j}^{\otimes 2}\exp(\hat{\theta}' X_{1j})\right\}^{-1}\hat{e}_{in},\\
\hat{d}_{in} &= \frac{1}{n}\sum_{j=1}^n \hat{\Lambda}_0(C_j)b_{in}(C_j) + \hat{\Lambda}_0(C_i) - \hat{\psi}.
\end{aligned}$$

Let α_0 and β_0 be the true values of α and β, respectively. Then, as we show in Appendix A, under some regularity conditions $n^{1/2}(\hat{\alpha} - \alpha_0, (\hat{\beta} - \beta_0)')'$ converges in distribution to a random normal variable with mean 0 and a covariance matrix that can be consistently estimated by $\hat{D}^{-1}\hat{\Sigma}\hat{D}^{-1}$, where $\hat{D} = n^{-1}\sum_{i=1}^{n} W_i X_{1i}^{\otimes 2}$, $\hat{\Sigma} = n^{-1}\sum_{i=1}^{n} \hat{\Phi}_i^{\otimes 2}$, and

$$\begin{aligned}\hat{\Phi}_i = & W_i X_{1i}\{\hat{\psi}^{-1}\exp(-\hat{\theta}' X_{1i})\bar{Y}_i - \hat{\alpha} - \hat{\beta}' X_i\} \\ & -\frac{1}{n}\sum_{j=1}^{n}\Big\{W_j X_{1j}\hat{\psi}^{-2}\exp(-\hat{\theta}' X_{1j})\bar{Y}_j\Big\}\hat{d}_{in} \\ & -\frac{1}{n}\sum_{j=1}^{n}\Big\{W_j X_{1j}^{\otimes 2}\hat{\psi}^{-1}\exp(-\hat{\theta}' X_{1j})\bar{Y}_j\Big\}\hat{f}_{in}.\end{aligned}$$

4. Model Diagnostics

For the checking of model (2.2), one has complete recurrent event data and can find some discussion and simple approaches in Huang and Wang (2004). Here we consider the assessment of model (2.1) and describe some graphical and numerical procedures for checking its adequacy. Let

$$\mathcal{A}(t) = \int_0^t \Big[\mu_0(u) + \frac{E\{g(Z)Z\}}{E(Z)}\Big] d\Lambda_0(u),$$

which can be estimated by

$$\hat{\mathcal{A}}(t) = \sum_{i=1}^{n}\int_0^t \frac{\{Y_i(u) - \hat{\beta}' X_i\}\Delta_i(u) dN_i(u)}{\sum_{i=1}^{n}\Delta_i(u)\exp(\hat{\theta}' X_{1i})},$$

where $\Delta_i(u) = I(C_i \geq u)$. For each i, following Lin et al. (2000) and Pan and Lin (2005), we define the residual

$$\hat{M}_i(t) = \int_0^t \Big[\Big\{Y_i(u) - \hat{\beta}' X_i\Big\}\Delta_i(u) dN_i(u) - \Delta_i(u)\exp(\hat{\theta}' X_{1i}) d\hat{\mathcal{A}}(u)\Big],$$

$i = 1, \ldots, n$. First we check the functional form for the kth component of X and plot $\hat{M}_i(t)$ against X_{ik}, where X_{ik} is the kth component of X_i. For a more formal procedure, let

$$\mathcal{F}_k(x) = n^{-1/2}\sum_{i=1}^{n} I(X_{ik} \leq x)\hat{M}_i(\tau),$$

the cumulative sum of $\hat{M}_i(t)$ over the values of X_{ik}. Let

$$S_0(t) = n^{-1}\sum_{i=1}^{n}\Delta_i(t)\exp(\hat{\theta}'X_{1i}),$$

$$S_k(t,x) = n^{-1}\sum_{i=1}^{n} I(X_{ik}\le x)\Delta_i(t)\exp(\hat{\theta}'X_{1i}),$$

$$B_1(t,x) = n^{-1}\sum_{i=1}^{n}\int_0^t \Big\{I(X_{ik}\le x) - \frac{S_k(u,x)}{S_0(u)}\Big\}X_i\Delta_i(u)dN_i(u),$$

$$B_2(t,x) = n^{-1}\sum_{i=1}^{n}\int_0^t \Big\{I(X_{ik}\le x) - \frac{S_k(u,x)}{S_0(u)}\Big\}X_i\Delta_i(u)\exp(\hat{\theta}'X_{1i})d\hat{\mathcal{A}}(u).$$

To apply the statistic $\mathcal{F}_k(x)$, we show in Appendix B that its null distribution can be approximated by the zero-mean Gaussian process

$$\tilde{\mathcal{F}}_k(x) = n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Big\{I(X_{ik}\le x) - \frac{S_k(u,x)}{S_0(u)}\Big\}d\hat{M}_i(u) - B_1(\tau,x)'(0_p, I_{p\times p})\hat{D}^{-1}n^{-1/2}\sum_{i=1}^{n}\hat{\Phi}_i - B_2(\tau,x)'n^{-1/2}\sum_{i=1}^{n}\hat{f}_{in}, \quad (4.1)$$

where 0_p is a p-dimensional vector of zeros, and $I_{p\times p}$ is a $p\times p$ identity matrix.

It is not possible to evaluate this distribution analytically because the limiting process of $\mathcal{F}_k(x)$ does not have an independent increments structure. For this, we propose using the simulation approach discussed in Cheng, Wei, and Ying (1997) and Lin et al. (2000). Let $(G_1,\ldots,G_n)$ be independent standard normal variables independent of the data. Then it can be shown, see Cheng, Wei, and Ying (1997) and Lin et al. (2000), that the distribution of the process $\mathcal{F}_k(x)$ can be approximated by that of the zero-mean Gaussian process

$$\hat{\mathcal{F}}_k(x) = n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Big\{I(X_{ik}\le x) - \frac{S_k(u,x)}{S_0(u)}\Big\}d\hat{M}_i(u)G_i - B_1(\tau,x)'(0_p, I_{p\times p})\hat{D}^{-1}n^{-1/2}\sum_{i=1}^{n}\hat{\Phi}_iG_i - B_2(\tau,x)'n^{-1/2}\sum_{i=1}^{n}\hat{f}_{in}G_i. \quad (4.2)$$

From (4.1) and (4.2), to approximate the distribution of $\mathcal{F}_k(x)$ one can obtain a large number of realizations from $\hat{\mathcal{F}}_k(x)$ by repeatedly generating the standard normal random sample $(G_1,\ldots,G_n)$ given the observed data. To assess the functional form of the jth component of covariates, one can plot a few realizations from $\hat{\mathcal{F}}_k(x)$ along with the observed $\mathcal{F}_k(x)$ to see if they can be regarded as arising from the same population. More formally, we can apply the supremum test statistic $\sup_x |\mathcal{F}_k(x)|$, where the p-value can be obtained by comparing the observed value of $\sup_x |\mathcal{F}_k(x)|$ to a large number of realizations of $\sup_x |\hat{\mathcal{F}}_k(x)|$.

An omnibus test for checking the overall fit of model (2.1) can be constructed from the process $\mathcal{F}_0(t,x) = n^{-1/2}\sum_{i=1}^n I(X_i \le x)\hat{M}_i(t)$, where the event $I(X_i \le x)$ means that each of the components of X_i is no larger than the corresponding component of x. As with $\mathcal{F}_k(x)$, we can show that the null distribution of $\mathcal{F}_0(t,x)$ can be approximated by that of the zero-mean Gaussian process $\hat{\mathcal{F}}_0(t,x)$, which is obtained from the expression (4.2) by replacing $I(X_{ik} \le x)$ with $I(X_i \le x)$, τ in the first integral with t, and $B_l(\tau,x)$ with $B_l(t,x)$ $(l = 1,2)$. An omnibus test statistic is then given by $\sup_{t,x} |\mathcal{F}_0(t,x)|$, based on which a p-value can be obtained as with $\sup_x |\mathcal{F}_k(x)|$.

5. Simulation Study

We conducted a simulation study to assess the estimation procedure proposed in the previous sections under different situations. In the study, the covariates X_i's were assumed to follow a Bernoulli distribution with success probability 0.5, or a normal distribution with mean zero and variance 0.25. To generate the simulated data, we first generated Z_i from the gamma distribution with mean 10 and variance 50, $g(Z_i) = \rho(Z_i - 10)/\sqrt{50}$, and the follow-up time C_i from the uniform distribution on $[\tau/2, \tau]$ with $\tau = 18$, respectively. Here ρ characterizes the relationship between the observation process and the longitudinal response process. When $\rho > 0$, the two processes are positively correlated; when $\rho = 0$, the two processes have no correlation given the covariates; when $\rho < 0$, the two processes are negatively correlated. Here, three situations with $\rho = -0.5$, 0, and 0.5 were considered.

For the observation process, we considered N_i as a homogeneous Poisson process with $\lambda_0(t) = \tau^{-1}$ or as a nonhomogeneous Poisson process with $\lambda_0(t) = (t+1)/\{\tau(\tau/2+1)\}$. For the first case, given X_i, Z_i, and C_i, K_i, the number of observation times for subject i, is Poisson with mean

$$\Lambda(C_i|X_i,Z_i) = Z_i\Lambda_0(C_i)\exp(X_i\gamma) = \frac{Z_iC_i\exp(\gamma X_i)}{\tau},$$

$i = 1,2,\ldots,n$, where $\gamma = 1$ was considered. The observation times $(T_{i1},\ldots,T_{i,K_i})$ were the order statistics of a random sample of size K_i from the uniform distribution over $(0, C_i)$.

For the second case, given X_i, Z_i, and C_i, K_i, the number of observation times for subject i, is Poisson with mean

$$\Lambda(C_i|X_i,Z_i) = Z_i\Lambda_0(C_i)\exp(\gamma X_i) = \frac{Z_i(C_i^2/2 + C_i)\exp(\gamma X_i)}{\tau(\tau/2+1)}.$$

The observation times $(T_{i1}, \ldots, T_{i,K_i})$ were the order statistics of a random sample of size K_i from the density function

$$\frac{t^2/2 + t}{C_i^2/2 + C_i} I(0 \leq t \leq C_i),$$

$i = 1, 2, \ldots, n$. Here $\gamma = 1$ was considered again.

For the response variable, it was assumed that

$$Y_i(t) = \mu_0(t) + \beta X_i + g(Z_i),$$

where $\mu_0(t) = 1 + t\sin(t)$. We took $\beta = -1, 0, 1$, representing different effects of the covariate X on the response variable. For each setting, we considered $n = 100$ and 200. All the results reported here were based on 1,000 Monte Carlo replications.

Tables 1-4 present the simulation results on estimation of β for the different situations. The tables include the biases (BIAS) given by the sample means of the proposed estimates of β minus the true values, the sample standard errors of the estimates (SSE) of $\hat{\beta}$, the means of the estimated standard errors (ESE) of $\hat{\beta}$, and the empirical 95% coverage probabilities (CP) for β. The results indicate that the biases of $\hat{\beta}$ are small and that the proposed variance estimation procedure provides reasonable estimates; empirical coverage probabilities indicate that the normal approximation seems to be appropriate. Note that β tends to be slightly underestimated for small sample sizes; this may be due to the use of the "borrow-strength estimation procedure" for estimation of θ. In addition, the variance seems underestimated; a possible reason is that the simulated data were generated from the joint model including random effects, and the estimating equation only involves the means of random effects. This does not seem to be a problem for large sample size. As seen in Tables 1-4, the estimated standard errors and the sample standard errors are quite close to each other, and the empirical 95% coverage probabilities are close to the nominal level.

We also carried out simulation studies to assess the robustness of the proposed approach compared with Sun, Sun, and Liu's approach. Here we took $g(Z_i) = \rho\log(Z_i/10 + 1) - E(\rho\log(Z_i/10 + 1))$. For generating X_i, Z_i, and C_i, we used the same setups as above. For generating the observation process, we let $\lambda_0(t) = 1/\tau$ and, given X_i, Z_i and C_i, K_i as Poisson with mean $\Lambda(C_i|X_i, Z_i)$ when $Z \leq 10$, and K_i as Poisson with mean 4 otherwise. To compare the performance with the estimators of Sun, Sun, and Liu (2007), we report the BIAS and SSE in Table 5 for the case that X_i is Bernoulli with success probability 0.5. Table 5 shows that our proposed estimators all had the smaller BIAS. Moreover, they were more efficient based on the SSEs. Our estimators perform well with the choice of g, but the estimators of Sun, Sun, and Liu (2007) had large bias

Table 1. Estimation of β with $\lambda_0(t) = \tau^{-1}$ when the $X_i \sim$ Bernoulli(0.5).

$\rho = 0.5$: Y and N are positively correlated						
β	1	0	-1	1	0	-1
		$n = 100$			$n = 200$	
BIAS	0.0062	0.0284	0.0066	0.0065	0.0230	0.0204
SSE	0.2589	0.2420	0.2429	0.1770	0.1749	0.1778
ESE	0.2419	0.2386	0.2437	0.1737	0.1702	0.1758
CP	0.9360	0.9420	0.9470	0.9410	0.9460	0.9560
$\rho = 0$: Y and N have no correlation						
β	1	0	-1	1	0	-1
		$n = 100$			$n = 200$	
BIAS	-0.0115	-0.0066	-0.0081	-0.0195	-0.0097	-0.0125
SSE	0.2143	0.2228	0.2276	0.1541	0.1559	0.1608
ESE	0.2137	0.2137	0.2279	0.1523	0.1526	0.1620
CP	0.9430	0.9400	0.9360	0.9520	0.9410	0.9540
$\rho = -0.5$: Y and N are negatively correlated						
β	1	0	-1	1	0	-1
		$n = 100$			$n = 200$	
BIAS	-0.0286	-0.0475	-0.0289	-0.0378	-0.0388	-0.0281
SSE	0.2532	0.2600	0.2797	0.1775	0.1767	0.1939
ESE	0.2403	0.2474	0.2613	0.1739	0.1779	0.1915
CP	0.9340	0.9340	0.9350	0.9430	0.9400	0.9390

because of the misspecification of the link function. The proposed method is robust, while Sun, Sun, and Liu's method is sensitive to the relationship between the longitudinal response process and the observation process.

6. An Application

To illustrate the proposed methodology, we consider a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Andrews and Herzberg (1985); Byar (1980); Sun and Wei (2000); Wellner and Zhang (2000); Zhang (2002)). In the study, the patients with superficial bladder tumors were randomly assigned to one of three treatment groups: placebo, thiotepa, or pyridoxine. During the study, many patients had multiple recurrences of the bladder tumors and all recurrences between visits were recorded and removed at clinical visits; the number of visits and visit time points varied greatly from patient to patient. At the beginning of the study, for each patient, two important baseline covariates were reported; the number of initial tumors and the size of the largest initial tumor. Following Sun and Wei (2000), we restrict our attention to the patients in the placebo (47) and the thiotepa (38) groups.

Table 2. Estimation of β with $\lambda_0(t) = (t+1)/\{\tau(\tau/2+1)\}$ when the $X_i \sim$ Bernoulli(0.5).

$\rho = 0.5$: Y and N are positively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0063	0.0136	0.0407	-0.0132	-0.0010	0.0398
SSE	0.3526	0.3455	0.3401	0.2543	0.2329	0.2352
ESE	0.3344	0.3213	0.3245	0.2451	0.2311	0.2310
CP	0.9340	0.9380	0.9340	0.9470	0.9540	0.9420
$\rho = 0$: Y and N have no correlation						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0608	-0.0329	-0.0172	-0.0609	-0.0317	-0.0011
SSE	0.3112	0.3133	0.3269	0.2227	0.2160	0.2298
ESE	0.3072	0.3025	0.3153	0.2212	0.2155	0.2248
CP	0.9410	0.9380	0.9380	0.9400	0.9540	0.9450
$\rho = -0.5$: Y and N are negatively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0945	-0.0832	-0.0598	-0.0852	-0.0800	-0.0458
SSE	0.3325	0.3328	0.3471	0.2364	0.2321	0.2561
ESE	0.3221	0.3247	0.3469	0.2339	0.2343	0.2508
CP	0.9260	0.9330	0.9320	0.9400	0.9430	0.9520

For the analysis, we took $Y_i(t)$ to be the logarithm of the number of observed tumors at time t, plus 1 to avoid 0, $i = 1, \ldots, 85$. We set the first component of X_i to 1 if the ith patient was given the thiotepa treatment and 0 otherwise, the second and the third components of X_i to the number of initial tumors and the size of the largest initial tumor of the ith patient, respectively, $i = 1, \ldots, 85$. The longitudinal process of the bladder tumors $Y_i(t)$ and the clinical visit process were described by models (2.1) and (2.2). The proposed application of the estimation procedure with $\eta_i = 1$ and $W_i = 1$ gave $\hat{\gamma} = (0.4808, -0.0358, 0.0156)'$ and $\hat{\beta} = (-0.7787, 0.1994, -0.0231)'$ with estimated standard errors $(0.128, 0.5767, 0.5287)'$ and $(0.2146, 0.0536, 0.0596)'$, and thus p-values $(0.0002, 0.9505, 0.9766)'$ and $(0.0003, 0.0002, 0.6987)'$, respectively. These results suggest that the thiotepa treatment significantly reduced the occurrence rate of the bladder tumors and the number of initial tumors has a significant positive effect on the tumor recurrence rate but no significant effect on the visit process. However, both the occurrence rate of the bladder tumors and the visit times did not seem to be significantly related to the size of the largest initial tumor. Sun, Sun, and Liu (2007) analyzed the same data and concluded that the thiotepa treatment had a significant effect in reducing the recurrence of bladder

Table 3. Estimation of β with $\lambda_0(t) = \tau^{-1}$ when the $X_i \sim N(0, 0.25)$.

$\rho = 0.5$: Y and N are positively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0105	0.0240	-0.0203	-0.0101	-0.0235	-0.0091
SSE	0.2825	0.2694	0.2907	0.2058	0.1906	0.2062
ESE	0.2639	0.2507	0.2855	0.1945	0.1782	0.1948
CP	0.9320	0.9490	0.9410	0.9530	0.9380	0.9430
$\rho = 0$: Y and N have no correlation						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0123	-0.0155	-0.0232	-0.0117	-0.0218	-0.0229
SSE	0.2427	0.2312	0.2686	0.1672	0.1606	0.1909
ESE	0.2218	0.2121	0.2514	0.1608	0.1544	0.1833
CP	0.9400	0.9340	0.9460	0.9500	0.9470	0.9480
$\rho = -0.5$: Y and N are negatively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0323	-0.0256	-0.0237	-0.0206	-0.0327	-0.0346
SSE	0.2708	0.2558	0.3060	0.1792	0.1897	0.2064
ESE	0.2377	0.2359	0.2766	0.1748	0.1733	0.2053
CP	0.9360	0.9320	0.9350	0.9480	0.9380	0.9510

tumors, but the initial number of bladder tumors had no significant effect in predicting the recurrence rate of the bladder tumor. There is a difference between our results and theirs. One possible reason is the misspecification of the relationship between the longitudinal response process and the observation process. As shown in Table 5, Sun, Sun, and Liu's approach is sensitive to the link function. Liang, Lu, and Ying (2009) also applied their method to the bladder tumor data, and their results showed that both the treatment indicator and the initial tumor number had significant effects on tumor recurrence rate. These results are consistent with those obtained by our proposed approach.

Consider the application of the model-checking procedures given in Section 4 to the data. Treating the three covariates separately, we found $\sup_x |\mathcal{F}_1(x)| = 1.5269$ with the p-value 0.377, $\sup_x |\mathcal{F}_2(x)| = 0.2230$ with the p-value 0.899, and $\sup_x |\mathcal{F}_3(x)| = 2.4113$ with the p-value 0.121. All three p-values suggest that we cannot reject model (2.1). We also checked the overall fit of model, and found $\sup_{t,x} |\mathcal{F}_0(t,x)| = 35.6916$ with the p-value 0.187, which yields the same conclusion.

7. Concluding Remarks

A key advantage of the proposed approach over existing methods for longitu-

Table 4. Estimation of β with $\lambda_0(t) = (t+1)/\{\tau(\tau/2+1)\}$ when the $X_i \sim N(0, 0.25)$.

$\rho = 0.5$: Y and N are positively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0103	-0.0337	-0.0233	-0.0052	-0.0066	-0.0343
SSE	0.4058	0.3280	0.3725	0.2818	0.2394	0.2671
ESE	0.3658	0.3144	0.3455	0.2685	0.2284	0.2481
CP	0.9390	0.9420	0.9320	0.9450	0.9440	0.9370
$\rho = 0$: Y and N have no correlation						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.03563	-0.0374	-0.0183	-0.0336	-0.0363	-0.0348
SSE	0.34630	0.3142	0.3593	0.2574	0.2163	0.2593
ESE	0.32520	0.2966	0.3391	0.2438	0.2132	0.2482
CP	0.93600	0.9400	0.9340	0.9420	0.9470	0.9370
$\rho = -0.5$: Y and N are negatively correlated						
β	1	0	-1	1	0	-1
	$n = 100$			$n = 200$		
BIAS	-0.0471	-0.0480	-0.0614	-0.0413	-0.0457	-0.0631
SSE	0.3677	0.3556	0.4010	0.2678	0.2464	0.2907
ESE	0.3350	0.3189	0.3797	0.2499	0.2348	0.2767
CP	0.9380	0.9320	0.9320	0.9320	0.9370	0.9360

Table 5. Simulation results of BIAS (SSE) of the proposed estimators and SSL's.

		$n = 100$		$n = 200$	
ρ	β	Proposed method	SSL's method	Proposed method	SSL's method
0.5	1	-0.0426(0.286)	-0.2229(0.3232)	-0.0449(0.2062)	-0.2302(0.2245)
	0	-0.0336(0.2791)	-0.2301(0.3116)	-0.0239(0.1957)	-0.2249(0.2203)
	-1	-0.0293(0.2634)	-0.2354(0.3078)	-0.0130(0.1959)	-0.2233(0.2237)
0	1	-0.0069(0.3033)	-0.1792(0.3111)	0.0056(0.2006)	-0.1768(0.2299)
	0	0.0124(0.2821)	-0.1842(0.3096)	0.0080(0.2025)	-0.1797(0.2463)
	-1	0.0251(0.2632)	-0.1763(0.3029)	0.0097(0.1863)	-0.1918(0.2255)
-0.5	1	0.0452(0.3043)	-0.1397(0.3181)	0.0329(0.2052)	-0.1412(0.2287)
	0	0.0551(0.2918)	-0.1444(0.3063)	0.0380(0.1983)	-0.1578(0.2282)
	-1	0.0720(0.2601)	-0.1282(0.2912)	0.0592(0.1902)	-0.1528(0.2236)

SSL's method stands for the one in Sun, Sun, and Liu (2007).

dinal data is that it allows the observation process to be related to the response process of interest through any unspecified link function of a latent variable. Another advantage is that the parameter estimates and the estimated covariance matrix do not involve estimation of the latent variables and the link function, while estimation of the latent variables are required by Sun, Sun, and Liu (2007), while the parametric distribution of the frailty variable and the link function are specified by Liang, Lu, and Ying (2009). In addition, our estimation procedure is more easily implemented. Simulations suggest that the proposed inference procedures perform well and an illustrative example is provided.

We have assumed that the follow-up process is independent of covariates for simplicity of presentation, and the proposed method can be generalized to the case where the censoring times may depend on covariates. For this case, following Sun and Wei (2000) assume that for subject i, the hazard function of C_i has the form

$$\lambda_C(t|X_i) = \lambda_{0C}(t)\exp(\phi' X_i), \tag{7.1}$$

where $\lambda_{0C}(t)$ is a completely unspecified baseline hazard function and ϕ is a p-dimensional vector of unknown regression parameters. To estimate β, motivated by (3.1) and $U(\alpha,\beta;\psi,\theta)$, consider the estimating function

$$\begin{aligned} &U_1(\alpha_1,\beta;\phi,\theta,S_0)\\ &\quad= \frac{1}{n}\sum_{i=1}^n W_i X_{1i}\left\{\exp(-\theta' X_{1i})\int_0^\tau \frac{Y_i(t)I(C_i\geq t)dN_i(t)}{\{S_0(t)\}^{\exp(\phi' X_i)}} - \alpha_1 - \beta' X_i\right\}\end{aligned}$$

for given ϕ, θ and S_0, where X_{1i}, W_i and θ are defined as before, $S_0(t) = \exp\{-\int_0^t \lambda_{0C}(s)ds\}$ denotes the baseline survival function of C, and

$$\alpha_1 = \int_0^\tau \Big[\mu_0(t) + \frac{E\{g(Z)Z\}}{E(Z)}\Big]d\Lambda_0(t).$$

In practice, ϕ, θ, and $S_0(t)$ are unknown, and we need to estimate them. Clearly, θ can be estimated by $\hat\theta$ as before. For estimation of ϕ and $S_0(t)$, we consider inference for the proportional hazards model (7.1) based on complete data. Then, following Kalbfleisch and Prentice (2002), one can estimate ϕ and $S_0(t)$ by the solution $\hat\phi$ to the estimating equation

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{X_i - \frac{\sum_{l=1}^n I(C_l\geq t)\exp(\phi' X_l)X_l}{\sum_{l=1}^n I(C_l\geq t)\exp(\phi' X_l)}\right\} dI(C_i\leq t) = 0,$$

and

$$\hat S_0(t) = \exp\left\{-\int_0^t \frac{\sum_{i=1}^n dI(C_i\leq s)}{\sum_{i=1}^n I(C_i\geq s)\exp(\hat\phi' X_i)}\right\},$$

respectively. Given $\hat\phi$, $\hat\theta$, and $\hat S_0(t)$, we propose to estimate α_1 and β by the solution $\hat\alpha_1$ and $\hat\beta$ to the estimating equation $U_1(\alpha_1,\beta;\hat\phi,\hat\theta,\hat S_0) = 0$. As before, one can show that $\hat\alpha_1$ and $\hat\beta$ are consistent and are asymptotically joint normal.

In the estimating equation approach, an important issue is how to choose the weights to improve the efficiency of estimation. The proposed estimation procedure involves the weight functions η_i and W_i. One can first choose the weight function η_i to improve the efficiency of the estimator of θ, and then choose the weight function W_i to improve the efficiency of the estimator of β. Further research is needed on this.

In the joint models, we assumed that the covariates are time-independent, but in some applications it would be desirable to develop estimation procedures that allow for both time-invariant and time-dependent covariates. For this, consider the joint models for the longitudinal response process $Y(t)$ and the observation process $N(t)$ as

$$E\{Y(t)|X(t),\xi\} = \mu_0(t) + \beta' X(t) + a'\xi + g(Z), \tag{7.2}$$

$$\lambda(t|X(t),\xi,Z) = \lambda_0(t) Z \exp(\gamma' X(t) + \varphi'\xi), \tag{7.3}$$

where $\lambda(t)$ is the intensity function of $N(t)$, $\lambda_0(t)$ is an unknown baseline intensity function, $X(t)$ is a vector of time-dependent covariates, ξ is a vector of time-independent covariates, Z is an unobserved random variable, and $g(\cdot)$ is an unknown link function. To estimate β and a in (7.2), we let

$$\bar{X}_i(\psi, S, \Lambda_0) = \int_0^\tau X_i(t) S(t) \frac{d\Lambda_0(t)}{\psi}$$

and propose the estimating function

$$\begin{aligned} &U_2(\alpha,\beta,a;\psi,\gamma,\theta,S,\Lambda_0) \\ &= \frac{1}{n}\sum_{i=1}^n W_i(\bar{X}_i', \xi_{1i}')' \left\{ \int_0^\tau \frac{Y_i(t) I(C_i \geq t) dN_i(t)}{\psi \exp(\gamma' X_i(t) + \theta'\xi_{1i})} - \alpha - \beta'\bar{X}_i - a'\xi_i \right\} \end{aligned}$$

for given ψ, γ, θ, S, and Λ_0, where $\xi_{1i} = (1, \xi_i')'$, $\theta = (\theta_1, \varphi')'$, W_i's are the weight functions, $S(t) = P(C \geq t)$, with $\psi = E\{\Lambda_0(C)\}$, $\theta_1 = \log(E(Z))$, and

$$\alpha = \psi^{-1} \int_0^\tau \Big[\mu_0(t) + \frac{E\{g(Z)Z\}}{E(Z)}\Big] P(C \geq t) d\Lambda_0(t).$$

In practice, ψ, γ, θ, $S(t)$, and $\Lambda_0(t)$ are unknown, and need to be estimated. Note that $S(t)$ can be estimated by its empirical survival function $S_n(t) = \sum_{i=1}^n I(C_i \geq t)/n$. For estimation of γ, θ, and $\Lambda_0(t)$, we consider inference for the intensity model (7.3) based on recurrent event data. Using the approach of Huang, Qin, and Wang (2010), one can obtain the estimators $\hat{\gamma}$, $\hat{\theta}$, and $\hat{\Lambda}_0(t)$, and ψ can be estimated by $\hat{\psi} = \sum_{i=1}^n \hat{\Lambda}_0(C_i)/n$. Given $\hat{\psi}$, $\hat{\gamma}$, $\hat{\theta}$, $S_n(t)$, and $\hat{\Lambda}_0(t)$, we propose to estimate α, β and a by the solution $\hat{\alpha}$, $\hat{\beta}$ and $\hat{a}$ to the estimating equation $U_2(\alpha,\beta,a;\hat{\psi},\hat{\gamma},\hat{\theta},S_n,\hat{\Lambda}_0) = 0$. As before, one can establish the consistency and the asymptotic normality of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{a}$. However, it seems not to be straightforward to generalize the proposed approach to the situation where the latent variables are also time-dependent, and further research is needed. It would also be of great interest to develop estimating procedures for the longitudinal regression model with time-varying coefficients when the longitudinal process depends on the observation process.

At (2.2), we assumed that a Poisson observation process $N(t)$. Further research is to replace (2.2) by the model

$$E\{N(t)|X,Z\} = \Lambda_0(t) Z \exp(\gamma' X),$$

where $\Lambda_0(t)$ is a completely unknown continuous baseline mean function. An estimation procedure needs to be developed for the joint mean models of a longitudinal response process and a general counting process.

Acknowledgements

The authors are grateful to two referees for valuable comments and suggestions that greatly improved this article. The research of the first author was supported in part by grants from the Research Grant Council of Hong Kong and The Hong Kong Polytechnic University. The research of the second author was supported in part by grants from the National Natural Science Foundation of China (NO. 10971015) and the Key Project of Chinese Ministry of Education (NO. 309007). The research of the third author was supported in part by grants from the National Natural Science Foundation of China Grants (No. 10731010, 10971015 and 10721101), the National Basic Research Program of China (973 Program) (No. 2007CB814902) and Key Laboratory of RCSDS, CAS (No.2008DP173182). .

Appendix A: Asymptotic Normality of $\hat{\alpha}$ and $\hat{\beta}$

To study the asymptotic distribution of the proposed estimates, we need the following regularity conditions.

(C1) $P(C \geq \tau, Z > 0) > 0$ and $E(Z^2) < \infty$.

(C2) X is bounded and $G(t) = E\{Z \exp(\gamma_0' X) I(C \geq t)\}$ is a continuous function for $t \in [0, \tau]$.

Let $R(t) = G(t)\Lambda_0(t)$, $H(t) = \int_0^t G(u) d\Lambda_0(u)$,

$$b_i(t) = \sum_{j=1}^{K_i} \left\{ \int_t^\tau \frac{I(T_{ij} \leq u \leq C_i) dH(u)}{R^2(u)} - \frac{I(t < T_{ij} \leq \tau)}{R(T_{ij})} \right\},$$

$$e_i(\theta) = -\int \frac{\eta x k b_i(c) dP_1(\eta, x, c, k)}{\Lambda_0(c)} + \eta_i X_{1i}[K_i\{\Lambda_0(C_i)\}^{-1} - \exp(\theta' X_{1i})],$$

where $P_1(\eta, x, c, k)$ is the joint probability measure of $(\eta_i, X_{1i}, C_i, K_i)$. Let

$$f_i(\theta) = \Big\{E\Big(-\frac{\partial e_i(\theta)}{\partial\theta}\Big)\Big\}^{-1} e_i(\theta).$$

Under conditions (C1) and (C2), it follows from Wang, Qin and Chiang (2001) that

$$n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} = n^{-1/2}\Lambda_0(t)\sum_{i=1}^{n} b_i(t) + o_p(1), \tag{A.1}$$

$$n^{1/2}(\hat{\theta} - \theta_0) = n^{-1/2}\sum_{i=1}^{n} f_i(\theta_0) + o_p(1), \tag{A.2}$$

where θ_0 is the true value of θ. Note that $n^{1/2}(\hat{\psi} - \psi_0) = I_1 + I_2$, where $I_1 = n^{-1/2}\sum_{i=1}^{n}\{\hat{\Lambda}_0(C_i) - \Lambda_0(C_i)\}$ and $I_2 = n^{-1/2}\sum_{i=1}^{n}\{\Lambda_0(C_i) - \psi_0\}$. Let $F_C(c)$ be the cumulative distribution function of C and $\hat{F}_C(c)$ be the corresponding empirical distribution based on $C_i, i = 1, \ldots, n$. It follows from (A.1) that

$$\begin{aligned} I_1 &= n^{1/2}\int\{\hat{\Lambda}_0(c) - \Lambda_0(c)\}d\hat{F}_C(c) \\ &= n^{1/2}\int\{\hat{\Lambda}_0(c) - \Lambda_0(c)\}dF_C(c) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\int \Lambda_0(c)b_i(c)dF_C(c) + o_p(1). \end{aligned}$$

Thus, we have

$$n^{1/2}(\hat{\psi} - \psi_0) = n^{-1/2}\sum_{i=1}^{n} d_i + o_p(1), \tag{A.3}$$

where

$$d_i = \int \Lambda_0(c)b_i(c)dF_C(c) + \{\Lambda_0(C_i) - \psi_0\}.$$

Note that

$$\begin{aligned} -\frac{\partial U(\alpha,\beta;\psi,\theta)}{\partial\psi} &= n^{-1}\sum_{i=1}^{n} W_i X_{1i}\psi^{-2}\exp(-\theta' X_{1i})\bar{Y}_i, \\ -\frac{\partial U(\beta,\alpha;\psi,\theta)}{\partial\theta} &= n^{-1}\sum_{i=1}^{n} W_i X_{1i}^{\otimes 2}\psi^{-1}\exp(-\theta' X_{1i})\bar{Y}_i. \end{aligned}$$

It follows from the Law of Large Numbers that $\partial U(\alpha,\beta;\psi,\theta)/\partial\psi|_{\psi=\psi_0,\theta=\theta_0}$ converges in probability to $-E\{W_i X_{1i}\psi_0^{-2}\exp(-\theta_0' X_{1i})\bar{Y}_i\}$, and $\partial U(\alpha,\beta;\psi,\theta)$ $/\partial\theta|_{\psi=\psi_0,\theta=\theta_0}$ converges in probability to $-E\{W_i X_{1i}^{\otimes 2}\psi_0^{-1}\exp(-\theta_0' X_{1i})\bar{Y}_i\}$.

Now, using (A.2), (A.3), and a Taylor series expansion, we have

$$\begin{aligned} n^{1/2}U(\alpha_0,\beta_0;\hat\psi,\hat\theta) &= n^{1/2}U(\alpha_0,\beta_0;\psi_0,\theta_0) \\ &\quad -E\{W_iX_{1i}\psi_0^{-2}\exp(-\theta_0'X_{1i})\bar Y_i\}n^{-1/2}\sum_{i=1}^n d_i \\ &\quad -E\{W_iX_{1i}^{\otimes 2}\psi_0^{-1}\exp(-\theta_0'X_{1i})\bar Y_i\}n^{-1/2}\sum_{i=1}^n f_i(\theta_0)+o_p(1), \end{aligned}$$

which converges in distribution to a normal random vector with mean 0 and covariance matrix $\Sigma = E(\Phi_i\Phi_i')$, where

$$\begin{aligned} \Phi_i &= W_iX_{1i}\{\psi_0^{-1}\exp(-\theta_0'X_{1i})\bar Y_i-\alpha_0-\beta_0X_i\} \\ &\quad -E\{W_iX_{1i}\psi_0^{-2}\exp(-\theta_0'X_{1i})\bar Y_i\}d_i \\ &\quad -E\{W_iX_{1i}^{\otimes 2}\psi_0^{-1}\exp(-\theta_0'X_{1i})\bar Y_i\}f_i(\theta_0). \end{aligned}$$

Note that $-(\partial\hat U(\alpha,\beta;\hat\psi,\hat\theta)/\partial\alpha, U(\alpha,\beta;\hat\psi,\hat\theta)/\partial\beta)$ converges in probability to $D = E\left\{W_iX_{1i}^{\otimes 2}\right\}$. Also note that a Taylor expansion of $U(\hat\alpha,\hat\beta;\hat\psi,\hat\theta)$ at $U(\alpha_0,\beta_0;\hat\psi,\hat\theta)$ yields

$$n^{1/2}\begin{pmatrix}\hat\alpha-\alpha_0\\ \hat\beta-\beta_0\end{pmatrix} = D^{-1}n^{1/2}U(\alpha_0,\beta_0,\hat\psi,\hat\theta)+o_p(1). \tag{A.4}$$

Therefore, $n^{1/2}(\hat\alpha-\alpha_0)$ and $n^{1/2}(\hat\beta-\beta_0)$ have an asymptotic joint normal distribution with mean 0 and covariance matrix $D^{-1}\Sigma D^{-1}$, which can be consistently estimated by $\hat D^{-1}\hat\Sigma\hat D^{-1}$.

Appendix B: Asymptotic Properties of $\mathcal{F}_k(x)$ and $\mathcal{F}_0(t,x)$

In the following, we only sketch the proof for the weak convergence of $\mathcal{F}_k(x)$ under models (2.1) and (2.2); the weak convergence of $\mathcal{F}_0(t,x)$ can be similarly derived. Assume that the limits of $S_k(t,x)$, $S_0(t)$, $B_1(t,x)$, and $B_2(t,x)$ exist and denote them by $s_k(t,x)$, $s_0(t)$, $b_1(t,x)$, and $b_2(t,x)$, respectively. Let

$$M_i(t)=\int_0^t\left[\Big\{Y_i(u)-\beta_0'X_i\Big\}\Delta_i(u)dN_i(u)-\Delta_i(u)\exp(\theta_0'X_{1i})d\mathcal{A}(u)\right].$$

For the weak convergence of $\mathcal{F}_k(x)$, using Lemma A.1 of Lin and Ying (2001) and the functional version of a Taylor expansion, we have

$$\begin{aligned} \mathcal{F}_k(x) &= n^{-1/2}\sum_{i=1}^n\int_0^\tau\Big\{I(X_{ik}\le x)-\frac{s_k(u,x)}{s_0(u)}\Big\}dM_i(u) \\ &\quad -b_1(\tau,x)'n^{1/2}(\hat\beta-\beta_0)-b_2(\tau,x)'n^{1/2}(\hat\theta-\theta_0)+o_p(1). \end{aligned} \tag{B.1}$$

The tightness of the first term on the right-hand side of (B.1) follows directly from the arguments in Appendix A.5 of Lin et al. (2000). The last two terms are also tight because $n^{1/2}(\hat{\beta}-\beta_0)$ and $n^{1/2}(\hat{\theta}-\theta_0)$ converge in distribution and $b_1(\tau,x)$ and $b_2(\tau,x)$ are some deterministic functions. It follows that $\mathcal{F}_k(x)$ is tight.

Based on (A.2) and (A.4), we can write $\mathcal{F}_k(x)$ as

$$\mathcal{F}_k(x) = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Big\{I(X_{ik}\le x) - \frac{s_k(u,x)}{s_0(u)}\Big\}dM_i(u)$$
$$-b_1(\tau,x)'(0_p, I_{p\times p})D^{-1}n^{-1/2}\sum_{i=1}^{n}\Phi_i - b_2(\tau,x)'n^{-1/2}\sum_{i=1}^{n}f_i(\theta_0) + o_p(1).$$

From the Multivariate Central Limit Theorem and tightness, $\mathcal{F}_k(x)$ converges weakly to a zero-mean Gaussian process which can be approximated by the zero-mean Gaussian process $\tilde{F}(t,z)$ given in (4.1).

References

Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker.* Springer-Verlag, New York

Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder tumors and other topics in urological oncology* (Edited by M. Pavane-Macaluso, P. H. Smith and F. Edsmyr), 363-370. Plenum, New York.

Cheng, S. C., Wei, L. J. and Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *J. Amer. Statist. Assoc.* **92**, 227-235.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *The Analysis of Longitudinal Data.* Oxford University Press, Oxford.

Huang, C.-Y., Qin, J. and Wang, M.-C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66**, 39-49.

Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99**, 1153-1165.

Liang, Y., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377-384.

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

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103-126.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics* **58**, 621-630.

Pan, Z. and Lin, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics* **61**, 1000-1009.

Sun, J., Park, D-H., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *J. Amer. Statist. Assoc.* **100**, 882-889.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *J. Amer. Statist. Assoc.* **102**, 1397-1406.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Roy. Statist. Soc. Ser. B* **62**, 293-302.

Wang, M. C., Qin, J., and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc.* **96**, 1057-1065.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779-814.

Welsh, A. H., Lin, X. and Carroll, R. J. (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel Methods. *J. Amer. Statist. Assoc.* **97**, 482-493.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39-48.

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

E-mail: xingqiu.zhao@polyu.edu.hk

School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, People's Republic of China,

E-mail: xweitong@bnu.edu.cn

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China.

E-mail: slq@amt.ac.cn

(Received October 2009; accepted November 2010)