Lifetime Data Anal (2009) 15:177–196
DOI 10.1007/s10985-008-9105-1

Semiparametric analysis of panel count data with correlated observation and follow-up times

Xin He · Xingwei Tong · Jianguo Sun

Received: 8 October 2007 / Accepted: 30 October 2008 / Published online: 10 December 2008

Abstract This paper discusses regression analysis of panel count data that often arise in longitudinal studies concerning occurrence rates of certain recurrent events. Panel count data mean that each study subject is observed only at discrete time points rather than under continuous observation. Furthermore, both observation and follow-up times can vary from subject to subject and may be correlated with the recurrent events. For inference, we propose some shared frailty models and estimating equations are developed for estimation of regression parameters. The proposed estimates are consistent and have asymptotically a normal distribution. The finite sample properties of the proposed estimates are investigated through simulation and an illustrative example from a cancer study is provided.

Keywords Estimating equation · Informative follow-up time · Informative observation times · Mean function model · Regression analysis

X. He
Division of Biostatistics, College of Public Health, The Ohio State University, B-116 Starling-Loving Hall, 320 West 10th Avenue, Columbus, OH 43210, USA
e-mail: xhe@cph.osu.edu

X. Tong (✉)
School of Mathematical Sciences, Beijing Normal University, Beijing 100875, People's Republic of China
e-mail: xweitong@bnu.edu.cn

J. Sun
Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA
e-mail: sunj@missouri.edu

1 Introduction

Panel count data usually arise in longitudinal follow-up studies that concern occurrence rates of certain recurrent events. In this situation, each study subject is observed only at discrete time points rather than under continuous observation and only the numbers of the events that occur between the observation times, not their occurrence times, are observed. Furthermore, both observation and follow-up times can vary from subject to subject. Areas that often produce panel count data include demographical studies, epidemiological studies, medical periodic follow-up studies and tumorgenicity experiments (Kalbfleisch and Lawless 1985; Thall and Lachin 1988).

For panel count data, three processes are involved. They are the underlying counting process that characterizes the recurrent process of interest, the process that governs observation times, and the process that determines follow-up times. If the three processes are independent completely or given covariates, a number of methods are available for the analysis of panel count data. For example, Kalbfleisch and Lawless (1985) considered the fitting of Markov model to panel count data. Sun and Kalbfleisch (1995) and Wellner and Zhang (2000) discussed estimation of mean function of the underlying counting process, while Sun and Wei (2000), Zhang (2002) and He et al. (2008) investigated regression analysis of panel count data.

In practice, the three processes may be correlated. For example, the follow-up times may be times to some terminal events related to the recurrent event of interest. Some recent references that discuss this for recurrent event data include Huang and Wang (2004), Liu et al. (2004), Huang and Liu (2007), Liu and Huang (2007), Rondeau et al. (2007) and Ye et al. (2007). Wang et al. (2001) also considered the same phenomenon and described a study of AIDS patients in which the recurrent and terminal events are hospitalization and death, respectively. Note that for recurrent event data, only the underlying counting process and the follow-up process are involved. Also one could face the same correlated problem in general longitudinal studies and for this, an extensive literature has been developed (De Gruttola and Tu 1994; Wulfsohn and Tsiatis 1997; Roy and Lin 2002; Song et al. 2002; Liu et al. 2007).

For panel count data, as mentioned above, one could have to deal with three related processes and the same could be true for longitudinal studies. For example, in the AIDS study discussed in Wang et al. (2001), suppose that one is interested in some symptoms related to AIDS such as CD4 counts or the time at which the patient's CD4 counts cross some threshold. Then the response process may be correlated with the observation process as well as the follow-up process. Lipsitz et al. (2002) presented a set of longitudinal data from a study of children with acute lymphoblastic leukemia which involves correlated response and observation processes. Huang et al. (2006) and Sun et al. (2007) discussed a set of panel count data that arose from a bladder cancer study in which the response process and the observation process are related. In this study, one may suspect the possible correlation among the follow-up process, the response process and the observation process as well. Therefore, in this paper, we consider situations where all three processes may be correlated with the main interest on the estimation of covariate effects on the response process after adjusting for the possible correlation among the three processes.

The remainder of this paper is organized as follows. Section 2 introduces notation and describes joint models for the three processes. To characterize the correlation, we employ some shared frailty models, a commonly used approach in both survival and longitudinal data analyses when a joint analysis is required. In Sect. 3, we consider estimation of regression parameters and for this, the estimating equation approach is applied. To implement the approach, a three-step estimation procedure is developed and the proposed estimates of regression parameters are consistent and have asymptotically a normal distribution. Section 4 presents some results obtained from a simulation study for assessing the proposed inference approach and an illustrative example is discussed in Sect. 5. Some concluding remarks and discussion are given in Sect. 6.

2 Notation and models

Consider a recurrent event study that consists of n independent subjects. Let $N_i(t)$ denote the number of occurrences of the recurrent event of interest before or at time t for subject i. Suppose that for subject i, there exists a vector of covariates denoted by x_i. Given x_i and two latent variables u_i and v_i, the mean function of $N_i(t)$ has the form

$$E\{N_i(t)|x_i, u_i, v_i\} = \mu_N(t)\ \exp(x_i'\beta_1 + u_i\beta_2 + v_i\beta_3)\,. \tag{1}$$

Here $\mu_N(t)$ is a completely unknown continuous baseline mean function and β_1, β_2 and β_3 are unknown regression parameters.

For subject i, suppose that $N_i(\cdot)$ is observed only at finite time points $T_{i1} < \cdots < T_{iK_i}$, where K_i denotes the potential number of observation times, $i = 1, \ldots, n$. That is, only the values of $N_i(t)$ at these observation times are known and we have panel count data on the $N_i(t)$'s. Also for subject i, suppose that there exist two follow-up times C_i^* and τ_i, where C_i^* may be related to $N_i(t)$ and the T_{il}'s and τ_i is independent of them. Assume that one only observes $C_i = \min(C_i^*, \tau_i)$ and $\delta_i = I(C_i = C_i^*)$ and thus $N_i(t)$ is observed only at these T_{il}'s with $T_{il} \le C_i, i = 1, \ldots, n$. Define $\tilde{N}_i(t) = H_i\{\min(t, C_i)\}$, where $H_i(t) = \sum_{l=1}^{K_i} I(T_{il} \le t), i = 1, \ldots, n$. Then $\tilde{N}_i(t)$ is a point process characterizing the ith subject's observation process and jumps only at the observation times.

In the following, we assume that given (x_i', u_i), $H_i(\cdot)$ is a non-homogeneous Poisson process with the intensity function

$$\lambda_{ih}(t) = \lambda_{0h}(t)\ \exp(x_i'\alpha_1 + u_i)\,. \tag{2}$$

In the model above, $\lambda_{0h}(t)$ is a completely unknown continuous baseline intensity function and α_1 denotes the vector of regression parameters. For the follow-up time C_i^*, it will be assumed that its hazard function is given by

$$\lambda_{ic}(t) = \lambda_{0c}(t)\ \exp(x_i'\gamma_1 + u_i\gamma_2 + v_i) \tag{3}$$

given x_i, u_i and v_i, where $\lambda_{0c}(t)$ denotes an unknown baseline hazard function and γ_1 and γ_2 are regression parameters.

Under models (1)–(3), it is clear that β_2, β_3 and γ_2 partly determine the correlation among the three processes. For instance, when $\gamma_2 = 0$, the observation process and the follow-up process are independent given (x_i', u_i). For subjects with the same (x_i', u_i), these two processes are positively correlated if $\gamma_2 > 0$ and they are negatively correlated if $\gamma_2 < 0$. Similarly, β_2 and β_3 measure the correlation between the response process and the observation process or the follow-up process, respectively, given (x_i', u_i, v_i). Therefore, one can estimate the covariate effects on the response process adjusting for possible correlation among the three processes. Note that there are no regression parameters associated with u_i and v_i in models (2) and (3) to avoid the identifiability issue and the difficulty to interpret the correlation between the response process and the observation process (measured by β_2) or the follow-up process (measured by β_3).

There exists a great deal of research on each of the three models (1)–(3) and their special cases individually. For example, model (3) without the latent variables is the well-known proportional hazards model (Kalbfleisch and Prentice 2002) and a number of methods have been developed for the same model with $\gamma_2 = 0$. Wang et al. (2001) and Huang and Wang (2004) considered a model similar to model (2) for recurrent event data. There also exists some limited work on the joint analysis of two of these models (Cheng and Wei 2000). In the following, we study the joint analysis of all three models together with the focus on estimation of regression parameters β_1 along with α_1 and γ_1. Let $\Lambda_{0h}(t) = \int_0^t \lambda_{0h}(s)ds$. We will assume that $\Lambda_{0h}(\tau) = 1$ for identifiability and $E(u_i|x_i) = E(u_i)$, where τ denotes the length of study. Also it will be assumed that $v_i \sim N(0, \sigma^2)$, where σ^2 is an unknown parameter.

3 Estimation of regression parameters

In this section, we consider estimation of β_1 along with other parameters. For this, note that if the latent effects u_i's and v_i's are known, then model (1) becomes the usual proportional means model and several methods such as that given in Cheng and Wei (2000) can be used. Unfortunately they are not known in practice. To deal with this, we borrow the idea used in Huang and Wang (2004) to first estimate or predict these unknown latent variables. For $i = 1, \ldots, n$, let $x_{1i}' = (x_i', u_i)$, $x_{2i}' = (x_i', u_i, v_i)$, $\beta' = (\beta_1', \beta_2, \beta_3)$, $\alpha' = (\alpha_1', 1, 0)$, and $\gamma' = (\gamma_1', \gamma_2)$. The proposed estimation procedure consists of the following three steps.

3.1 Estimation of model (2)

To estimate β_1, we first consider inference about model (2), for which we have recurrent event data. Let $K_i^* = \tilde{N}_i(C_i)$, the total number of observations on subject i, $i = 1, \ldots, n$. Also let the s_j's denote the ordered and distinct time points of all the observation times $\{T_{il}\}$, d_j be the number of the observation times equal to s_j, and n_j be the number of the observation times satisfying $T_{il} \leq s_j \leq C_i$ among all subjects. Define $x_{3i}' = (x_i', 1)$, $\alpha_*' = (\alpha_1', \alpha_2) = (\alpha_1', E(u_i))$. Then following Huang and Wang (2004), one can first estimate $\Lambda_{0h}(t)$ and α_* by

$$\widehat{\Lambda}_{0h}(t) = \prod_{s_l > t} \left(1 - \frac{d_l}{n_l}\right)$$

and the estimating equation

$$\sum_{i=1}^{n} w_i\, x_{3i} \left\{ K_i^* \, \widehat{\Lambda}_{0h}^{-1}(C_i) - \exp(\alpha_*' x_{3i}) \right\} = 0, \tag{4}$$

respectively. In Eq. 4, the w_i's are some weights that could depend on x_i, C_i and Λ_{0h}. A key fact used in deriving the above estimating equation is that conditional on (x_i', C_i, u_i, K_i^*), the observation times $\{T_{i1}, \ldots, T_{iK_i^*}\}$ are the order statistics of a simple random sample of size K_i^* from the density function

$$\frac{\lambda_{0h}(t) \exp(\alpha_1' x_i + u_i)}{\Lambda_{0h}(C_i)\, \exp(\alpha_1' x_i + u_i)}\, I(0 \le t \le C_i) = \frac{\lambda_{0h}(t)}{\Lambda_{0h}(C_i)}\, I(0 \le t \le C_i)\,.$$

Let $\hat{\alpha}_*' = (\hat{\alpha}_1', \hat{\alpha}_2)$ denote the estimate of α_*' given by Eq. 4. Note that given (x_i', C_i, u_i), the expected value of K_i^* is equal to $\Lambda_{0h}(C_i)\, \exp(\alpha_1' x_i + u_i)$. Thus it is natural to predict u_i by

$$\hat{u}_i = \log \left\{ \frac{K_i^*}{\hat{\Lambda}_{0h}(C_i)\, e^{\hat{\alpha}_1' x_i}} \right\}. \tag{5}$$

3.2 Estimation of model (3)

In this subsection, we discuss estimation of model (3). For this, let $O = (O_1, \ldots, O_n)$, where $O_i' = (C_i, \delta_i, x_i', u_i)$ denotes the observed data on subject i assuming that u_i is known. Also let $c_1 < \cdots < c_k$ denote the ordered observed failure times and assume that we can write $\Lambda_{0c}(t)$ as

$$\Lambda_{0c}(t) = \sum_{j=1}^{k} a_j\, I(t \ge c_j)\,,$$

where $a' = (a_1, \ldots, a_k)$ is a vector of unknown parameters. Define $\theta = (a', \gamma', \sigma^2)'$. Then the full likelihood function has the form

$$L(\theta) = \prod_{i=1}^{n} \{\lambda_{0c}(C_i) \exp(x_{1i}'\gamma + v_i)\}^{\delta_i} \exp\{-\Lambda_{0c}(C_i) \exp(x_{1i}'\gamma + v_i)\}\, \phi(v_i; \sigma)$$

based on the pseudo complete data O and the v_i's, where $\phi(\cdot; \sigma)$ denotes the density function of $N(0, \sigma^2)$.

To maximize $L(\theta)$ with respect to θ, we propose to replace u_i in $L(\theta)$ by its prediction given in (5) and then use the EM algorithm to deal with the latent variables v_i's as usual. To implement the EM algorithm, we first consider the E-step, which computes the conditional expectation of the log likelihood function given the current estimate of θ and the observed data O. To this end, note that the log likelihood function can be written as

$$l(\theta) = \sum_{i=1}^{n} \left\{ \delta_i \left[\log\{\lambda_{0c}(C_i)\} + x_{1i}'\gamma + v_i \right] - \Lambda_{0c}(C_i)\exp(x_{1i}'\gamma + v_i) + \log\phi(v_i;\sigma) \right\}$$
$$= \sum_{i=1}^{n} \delta_i \left[\log\{\lambda_{0c}(C_i)\} + x_{1i}'\gamma \right] + \sum_{i=1}^{n} g(v_i;\theta),$$

where

$$g(v_i;\theta) = \delta_i v_i - \Lambda_{0c}(C_i)\exp(x_{1i}'\gamma + v_i) + \log\phi(v_i;\sigma).$$

To calculate $E\{l(\theta)|O,\theta^{(m)}\}$, one needs to calculate

$$E_i\{g(v_i;\theta)|O_i,\theta^{(m)}\} = \int g(v_i;\theta)\, f(v_i|O_i,\theta^{(m)})\, dv_i \,,$$

where $\theta^{(m)}$ denotes the current estimate of θ and

$$f(v_i|O_i,\theta) = \frac{\exp(\delta_i v_i)\exp\{-\Lambda_{0c}(C_i)\exp(x_{1i}'\gamma + v_i)\}\phi(v_i;\sigma)}{\int \exp(\delta_i v_i)\exp\{-\Lambda_{0c}(C_i)\exp(x_{1i}'\gamma + v_i)\}\phi(v_i;\sigma)dv_i},$$

the conditional density of v_i given O_i and θ. It is apparent that this integration has no closed form. For this, with $\theta = \theta^{(m)}$, let $\{v_i^{(l)}; i = 1,\ldots,n, l = 1,\ldots,L\}$ be L independent and identically distributed samples from $N(0,\{\sigma^{(m)}\}^2)$, where L is sufficiently large. Then one can approximate $E_i\{g(v_i;\theta)|O_i,\theta^{(m)}\}$ by

$$\hat{E}_i\{g(v_i;\theta)|O_i,\theta^{(m)}\}$$
$$= \frac{\sum_{l=1}^{L} g(v_i^{(l)};\theta)\exp(\delta_i v_i^{(l)})\exp\{-\Lambda_{0c}^{(m)}(C_i)\exp(x_{1i}'\gamma^{(m)} + v_i^{(l)})\}}{\sum_{l=1}^{L} \exp(\delta_i v_i^{(l)})\exp\{-\Lambda_{0c}^{(m)}(C_i)\exp(x_{1i}'\gamma^{(m)} + v_i^{(l)})\}}. \quad (6)$$

Now we consider the M-step of the EM algorithm, which maximizes $E\{l(\theta)|O,\theta^{(m)}\}$ with respect to θ. For this, by taking its derivatives with respect to θ and setting the derivatives equal to zero, we obtain the following equations

$$a_j^{(m+1)} = \left[\sum_{i=1}^{n} E_i\left\{\exp(x_{1i}'\gamma + v_i) I(C_i \geq c_j)\right\} \right]^{-1}, \quad (7)$$

for $j = 1, .., k$, $\sigma^{(m+1)} = \{n^{-1} \sum_{i=1}^{n} E_i(v_i^2)\}^{1/2}$, and

$$\sum_{i=1}^{n} E_i \left[x_{1i} \left\{ \delta_i - \Lambda_{0c}(C_i) \exp(x_{1i}'\gamma + v_i) \right\} \right] = 0 \tag{8}$$

for the updated estimate $\theta^{(m+1)}$ of θ. In practice, we propose to obtain the $a_j^{(m+1)}$ and thus $\Lambda_{0c}^{(m+1)}$ first by using (7) with letting $\theta = \theta^{(m)}$. Then by replacing Λ_{0c} with $\Lambda_{0c}^{(m+1)}$, one can solve (8) to get $\gamma^{(m+1)}$ and $\{\sigma^{(m+1)}\}^2$. Finally, given the estimate $\hat{\theta}$ of θ, one could calculate the conditional expectation of v_i given O_i as or predict v_i by

$$\hat{v}_i = \hat{E}_i(v_i | O_i, \hat{\theta}), \tag{9}$$

which can be approximated by (6).

3.3 Estimation of β_1

Now we are ready to estimate β_1 or β in model (1). For this, define $Y_i(t) = I(t \leq C_i)$ and

$$S_j(\beta; t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \exp\{x_{2i}'(\beta + \alpha)\} \, x_{2i}^{\otimes j},$$

for $j = 0, 1, 2$, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = a\,a'$ for a vector a. Note that if all the u_i's and v_i's are known, following Cheng and Wei (2000), one can estimate β using the estimating function

$$U(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ x_{2i} - \frac{S_1(\beta; t)}{S_0(\beta; t)} \right\} N_i(t) d\tilde{N}_i(t).$$

Motivated by this, we propose to estimate β based on the following estimating function

$$\hat{U}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{x}_{2i} - \frac{\hat{S}_1(\beta; t)}{\hat{S}_0(\beta; t)} \right\} N_i(t) d\tilde{N}_i(t). \tag{10}$$

Here $\hat{x}_{2i} = (x_i', \hat{u}_i, \hat{v}_i)'$ with the $\hat{u}_i$ and $\hat{v}_i$ given by (5) and (9), respectively, and $\hat{S}_j(\beta; t)$ denotes $S_j(\beta; t)$ with the x_{2i}'s and α replaced by the $\hat{x}_{2i}$'s and $\hat{\alpha} = (\hat{\alpha}_1', 1, 0)'$, respectively.

Define the estimate $\hat{\beta}$ of β as the solution to $\hat{U}(\beta) = 0$. Note that $\hat{x}_{2i}$ is independent of β and we have

$$\frac{\partial \hat{U}(\beta)}{\partial \beta} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{0}^{\tau} \left\{ \frac{\hat{S}_2(\beta; t)\hat{S}_0(\beta; t) - \hat{S}_1(\beta; t)\hat{S}_1(\beta; t)'}{\hat{S}_0^2(\beta; t)} \right\} N_i(t)\, d\tilde{N}_i(t),$$

which is strictly negative. Thus $\hat{\beta}$ is unique. In the Appendix, we show that $\hat{\beta}$ is consistent and $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a normal random vector with mean zero and covariance matrix given in the Appendix.

For estimation of the covariance matrix of $\hat{\beta}_1$ or $\hat{\beta}$, we propose to use the following simple bootstrap procedure. Let B denote a prespecified positive integer. For each b, where $1 \le b \le B$, draw a simple random sample of size n,

$$D^{(b)} = \left\{ T_{i1}^{(b)}, \ldots, T_{iK_i^*}^{(b)}, N_i^{(b)}(T_{i1}^{(b)}), \ldots, N_i^{(b)}(T_{iK_i^*}^{(b)}), C_i^{(b)}, \delta_i^{(b)}, x_i^{(b)\prime};\ i = 1, \ldots, n \right\},$$

with replacement from the observed data

$$D = \left\{ T_{i1}, \ldots, T_{iK_i^*}, N_i(T_{i1}), \ldots, N_i(T_{iK_i^*}), C_i, \delta_i, x_i';\ i = 1, \ldots, n \right\}.$$

Let $\hat{\beta}^{(b)}$ be the proposed estimate of β based on the data set $D^{(b)}$ defined above. Then a natural estimate of the covariance matrix of $\hat{\beta}$ is given by

$$\hat{\Sigma} = \frac{1}{B-1} \sum_{b=1}^{B} \left\{ \hat{\beta}^{(b)} - \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}^{(b)} \right\}^{\otimes 2}.$$

In the procedure given above, one needs to choose L and B. In general, for a practical problem, one may start with some reasonable values and then increase them until the results are stable. For example, it is common to choose $L = 200$ and $B = 100$. For simulation studies, if using enough replications, one may actually only need to use small values for them to save the computational effort. More comments are given in the next section.

4 Numerical results

We conducted a simulation study to assess the performance of the estimation procedure proposed in the previous sections under different situations with the focus on estimation of β_1. Note that models (2) and (3) and the estimation of the regression parameters in them have been investigated by other authors. In the study, the covariate x_i's were assumed to follow a Bernoulli distribution with success probability 0.5. To generate the simulated data, we first generated the $u_i^* = \exp(u_i)$ and v_i from the gamma distribution with mean 10 and variance 50 and the normal distribution with mean 0 and variance $\sigma^2 = 0.25$, respectively. Given x_i, u_i and v_i, C_i^* was generated under model (3) with $\lambda_{0c}(t) = 0.002$ and it was assumed that $\tau_i = \tau = 18$.

For the observation process, we assumed that H_i follows the homogeneous Poisson process with $\lambda_{0h}(t) = \tau^{-1}$. Then given x_i and u_i, K_i^*, the number of real observation times for subject i, follows the Poisson distribution with mean

$$\Lambda_{ih}(C_i|x_i, u_i) = \Lambda_{0h}(C_i)\exp(x_i\alpha_1 + u_i) = \frac{C_i \exp(x_i\alpha_1 + u_i)}{\tau},$$

$i = 1, 2, \ldots, n$. Furthermore, the observation times $(T_{i1}, \ldots, T_{iK_i^*})$ are the order statistics of a random sample of size K_i^* from the uniform distribution over $(0, C_i)$. Given K_i^* and $(T_{i1}, \ldots, T_{iK_i^*})$, we generated $N_i(T_{ij})$ using the formula

$$\begin{aligned} N_i(T_{ij}) = N_i^*[\lambda_N(T_{i1})] + N_i^*[\lambda_N(T_{i2}) - \lambda_N(T_{i1})] \\ + \cdots + N_i^*[\lambda_N(T_{ij}) - \lambda_N(T_{i\,j-1})] \end{aligned}$$

for $j = 1, \ldots, K_i^*$ and $i = 1, \ldots, n$. Here $N_i^*[\lambda_N(t)]$ denotes the random number generated from the Poisson distribution with mean

$$t\ \exp(x_i\beta_1 + u_i\beta_2 + v_i\beta_3)\,.$$

The results given below are based on $n = 100$ or 200, and 1000 replications.

Table 1 presents the simulation results on estimation of β_1 for the situations where $\beta_1 = -2, -1, 0, 1, 2$ along with $\beta_2 = \beta_3 = 0$ and $\alpha_1 = \gamma_1 = \gamma_2 = 1$. The table includes the averages of proposed estimates of β_1 based on the simulated data, the sample standard deviations of the estimates (SSD), the means of the bootstrap standard deviation estimates (BSD), and the empirical 95% coverage probabilities (CP) for β_1. Table 2 gives the estimation results for the same situations as in Table 1 except that

Table 1 Estimation of β_1 with $\beta_2 = \beta_3 = 0$, $L = 50$ and $B = 20$

	True β_1				
	−2	−1	0	1	2
$n = 100$					
$\hat{\beta}_1$	−2.0159	−1.0019	0.0027	0.9966	2.0071
SSD	0.1843	0.1224	0.0967	0.0807	0.0778
BSD	0.1847	0.1287	0.1001	0.0876	0.0837
CP	0.930	0.943	0.937	0.946	0.945
$n = 200$					
$\hat{\beta}_1$	−2.0143	−1.0045	−0.0013	1.0018	2.0009
SSD	0.1273	0.0877	0.0660	0.0601	0.0539
BSD	0.1282	0.0900	0.0708	0.0624	0.0583
CP	0.930	0.942	0.957	0.948	0.947

Table 2 Estimation of β_1 with $\beta_2 = \beta_3 = 0.2$, $L = 50$ and $B = 20$

	True β_1				
	−2	−1	0	1	2
$n = 100$					
$\hat{\beta}_1$	−2.0135	−1.0037	−0.0097	0.9936	1.9925
SSD	0.1700	0.1264	0.1107	0.1043	0.1076
BSD	0.1644	0.1248	0.1096	0.1023	0.1036
CP	0.920	0.933	0.944	0.934	0.942
$n = 200$					
$\hat{\beta}_1$	−2.0097	−1.0051	−0.0054	0.9938	1.9980
SSD	0.1334	0.0941	0.0895	0.0861	0.0890
BSD	0.1242	0.0950	0.0847	0.0838	0.0827
CP	0.938	0.946	0.940	0.939	0.940

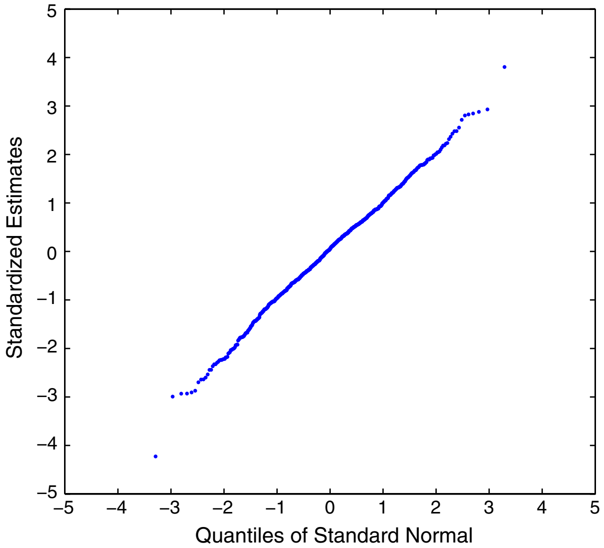

Fig. 1 Quantile plot with $\beta_1 = 0$, $\beta_2 = 0$ and $\beta_3 = 0$

$\beta_2 = \beta_3 = 0.2$. These results indicate that the estimate $\hat{\beta}_1$ seems to be unbiased and the bootstrap variance estimation procedure provides reasonable estimates. Also the results on the empirical coverage probabilities indicate that the normal approximation seems to be appropriate.

As mentioned above, in the study, we have focused only on β_1 to save the space and actually the results for other parameters are similar to those given above. For example, for the setups considered in Table 1 with $\beta_1 = 0$, we obtained $\hat{\beta}_2 = 0.0016$,

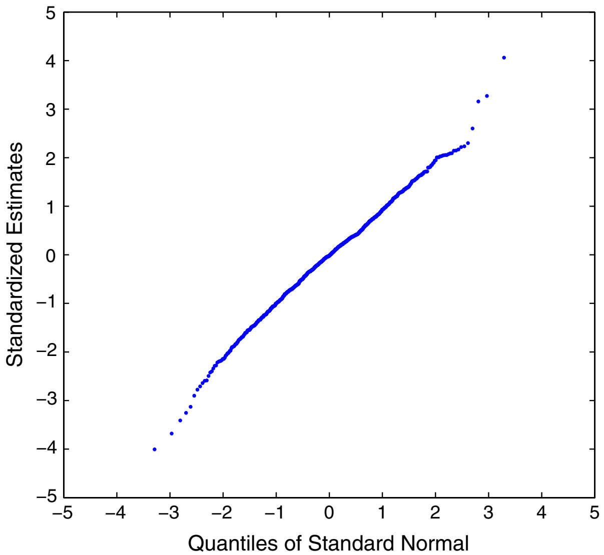

Fig. 2 Quantile plot with $\beta_1 = 1$, $\beta_2 = 0$ and $\beta_3 = 0$

Table 3 Estimation of β_1 with $\beta_2 = \beta_3 = 0$, $L = 200$ and $B = 100$

	True β_1				
	−2	−1	0	1	2
$n = 100$					
$\hat{\beta}_1$	−2.0169	−1.0037	0.0046	1.0011	2.0039
SSD	0.1759	0.1212	0.0923	0.0845	0.0791
BSD	0.1840	0.1279	0.1004	0.0878	0.0834
CP	0.947	0.949	0.956	0.943	0.957

$\hat{\beta}_3 = 0.0031, \hat{\alpha}_1 = 0.9652, \hat{\gamma}_1 = 0.9848$ and $\hat{\gamma}_2 = 1.0506$ with $n = 100$, and $\hat{\beta}_2 = -0.0002, \hat{\beta}_3 = -0.0009, \hat{\alpha}_1 = 1.0034, \hat{\gamma}_1 = 1.0083$ and $\hat{\gamma}_2 = 1.0469$ with $n = 200$ based on the simulated data. To assess the performance of the normal approximation to the finite-sample distribution of the estimate of β_1, we studied the quantile plots of the standardized estimates of β_1. Figures 1 and 2 display such plots corresponding to the situations where $n = 100$, $\beta_1 = 0$ or 1 with $\beta_2 = \beta_3 = 0$. These figures indicate that the normal approximation seems reasonable. Similar plots were obtained for other setups.

Note that in Tables 1 and 2, we used $L = 50$ and $B = 20$. As mentioned before, in practice, one should use large values for them. For the simulation study here, however, these values seem to be sufficient. To give an example, Table 3 gives the results obtained for the proposed estimate of β_1 under the same setup as that for Table 1 except that $L = 200$, $B = 100$ and $n = 100$. Similar simulation results were obtained for other setups with larger values of L and B.

5 An application

In this section, we illustrate the proposed methodology using the data set from a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Byar 1980; Andrews and Herzberg 1985; Wellner and Zhang 1998; Sun and Wei 2000). In the study, the patients with superficial bladder tumors were randomly assigned to one of three treatment groups: placebo, thiotepa and pyridoxine. During the study, many patients had multiple recurrences of the bladder tumors and all recurrent tumors between the visits were recorded and removed at their clinical visits. Both 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 and they are 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 thiotepa (38) groups. For these two groups, the average numbers of visits were 8.66 and 13.50 and the median follow-up times were 30 months and 32.5 months, respectively. Also the average numbers of bladder tumor recurrence were 39.81 and 17.03, respectively.

For the analysis, define the first component of x_i to be equal to 1 if the ith patient was given the thiotepa treatment and 0 otherwise. Also define the second and third components of x_i to be the number of initial tumors and the size of the largest initial tumor of the patient, respectively. Assume that the occurrence process of the bladder tumors, the clinical visit process and the follow-up process can be described by models (1), (2) and (3), respectively. The application of the estimation procedure with $L = 200$ and $B = 1000$ proposed in the previous sections gave $\hat{\beta}_1 = (-1.8483, 0.1996, 0.0015)'$ with the estimated standard errors of $(0.6879, 0.3181, 0.3562)'$. These results suggest that the thiotepa treatment significantly reduced the occurrence rate of the bladder tumors. However, the occurrence rate of the bladder tumors does not seem to be significantly related with the number of initial tumors and the size of the largest initial tumor. Note that for the analysis here, we also tried larger values of L and B and obtained similar results.

For comparison, we noticed that Sun and Wei (2000) assumed that the three processes involved are independent of each other given the covariates and estimated the effects of three covariates as $(-2.0249, 0.6620, -0.1229)'$ with the estimated standard errors of $(0.4500, 0.2133, 0.2035)'$. It can be seen that the results from the two methods are similar but the approach that took into account the possible correlation among the three processes gave smaller estimated effects. In other words, without taking into account the correlation, one could overestimate the treatment or covariate effects. One possible reason for this is that part of the estimated effects given by the approach assuming the independence may be due to the correlation of the three processes. Huang et al. (2006) studied a similar problem and gave a similar conclusion.

6 Concluding remarks

In this article, we considered regression analysis of panel count data when all three processes involved may be related and for the purpose, some shared frailty models were proposed. For inference, an estimating equation approach and an EM algorithm were developed for estimation of regression parameters representing covariate effects. A key advantage of the proposed approach over existing methods for panel count data is that it allows both the observation process and the follow-up process to be related with the response process of interest. In general, it may be hard to have enough evidence to verify or assess the independence assumption or the existence of the correlation. But as mentioned before, this could happen quite often in practice.

In the preceding sections, our focus has been on estimation of regression parameter β_1. Sometimes, one may be interested in estimating the baseline mean function $\mu_N(t)$ in model (1). For this, one could develop an estimate following the one given in Thall and Lachin (1988) or others by treating u_i and v_i to be known and replaced with $\hat{u}_i$ and $\hat{v}_i$ given in (5) and (9), respectively.

For the estimation of regression parameters, we developed an EM algorithm, which is complicated in computation. As an alternative, one may apply the approach given in Louis (1982). However, this alternative is also very computationally intensive and its detailed investigation and comparison with the approach given here are beyond the scope of this paper. It is worth to mention that in this paper, we only considered time-independent covariates. For the cases with time-dependent covariates, one can still use the estimating function (10) with respect to model (1) but may need different estimation processes with respect to models (2) and (3).

Note that the estimating Eq. 4 involves a weight function w_i which could depend on x_i, C_i and Λ_{0h} but no procedure was provided for its selection. As a direction for future research, it would be useful to investigate how to choose this weight function to obtain efficient or optimal estimates. Similarly, one could apply some weight functions to the estimating function (10) and ask the same question. Another direction for future research is to conduct a sensitivity analysis with respect to the frailty and to develop some goodness-of-fit tests for the proposed models.

Acknowledgements We wish to thank the Associate Editor and two reviewers for their many thoughtful and helpful comments and suggestions, which greatly improved the paper. This project was supported in part by NIH grant OH008649 (He) and Scientific Research Foundation for Returned Scholars, Ministry of Education of China (Tong).

Appendix

Proof of the asymptotic properties of $\hat{\beta}$

In this appendix, we prove the consistency and normality of $\hat{\beta}$. Let N_i, $\tilde{N}_i$, T_{il}, K_i^*, x_i, x_{1i}, x_{2i}, x_{3i}, u_i, v_i, w_i and C_i be defined as in the previous sections, and K^*, X, X_1, X_2, X_3, U, V, W and C denote the underlying random variables of the K_i^*'s, x_i's, x_{1i}'s, x_{2i}'s, x_{3i}'s, u_i's, v_i's, w_i's and C_i's, respectively. In order to study the asymptotic properties of $\hat{\beta}$, we first impose the following regular conditions that are similar to those given in Huang and Wang (2004).

(a) $P(C \geq \tau, \exp(U) > 0) > 0$;
(b) X is uniformly bounded;
(c) The variance of $\exp(U)$ is bounded and there exists a positive small constant $\epsilon > 0$ such that $\exp(U) > \epsilon$ almost surely;
(d) $G(s) = E\{\exp(U)I(C \geq s)\}$ is continuous for $s \in [0, \tau]$.

Following the procedures in Fan and Li (2002), to achieve the good statistical properties of the estimators $\hat{\gamma}$ and $\hat{\Lambda}_{0c}(t)$ in model (3), the following two conditions are also assumed to be satisfied.

(e) $\Lambda_{0c}(\tau) < \infty$.
(f) $(\beta', \gamma', \alpha')'$ is an inner point of a compact subset in R^{3p+5};

Let $e_k, k = 1, \ldots, p+2$, be a $(p+2)-$dimensional vector whose elements are all zero except its kth entry being equal to one. Using the same notation as Wang et al. (2001), define $Q(u) = \int_0^u G(v)d\Lambda_{0h}(v)$, $R(u) = G(u)\Lambda_{0h}(u)$,

$$b_{ih}(t) = \sum_{j=1}^{K_i^*}\left\{\int_t^\tau \frac{I(T_{ij} \leq u \leq C_i)dQ(u)}{R^2(u)} - \frac{I(t \leq T_{ij} \leq \tau)}{R(T_{ij})}\right\}$$

and

$$f_i = \int \frac{WX_3K^*b_{ih}(C)dP(W, X_3, K^*, C)}{\Lambda_{0h}(C)} + w_i x_{3i}\{K_i^*\Lambda_{0h}^{-1}(C_i) - \exp(x_{3i}'\alpha_*)\},$$

where $P(\cdot)$ denotes the joint distribution of the underlying variables. Then we have

$$\widehat{\alpha}_1 - \alpha_1 = \frac{1}{n}\sum_{i=1}^n f_{ih} + o_p(n^{-1/2}), \tag{A.1}$$

where f_{ih} is the vector function $E\{-\partial f_i/\partial\alpha_*\}^{-1} f_i$ without the last entry, and

$$\widehat{\Lambda}_{0h}(t) - \Lambda_{0h}(t) = \frac{1}{n}\Lambda_{0h}(t)\sum_{i=1}^n b_{ih}(t) + o_p(n^{-1/2}), \quad t \leq \tau \tag{A.2}$$

(Wang et al. 2001; Huang and Wang 2004).

In this paper, we assume that the censoring time C follows the Cox's frailty model and that the latent variable V follows the normal distribution. Fan and Li (2002) discussed this issue where the frailty variable $\exp(V)$ follows the Gamma distribution. Under the conditions (a)–(f) listed above, the conditions A–D in Fan and Li (2002) are all satisfied. Using the same arguments, there exist functions $b_{ic}(\cdot)$ and independent and identically distributed random variables $f_{ic}, i = 1, \ldots, n$, such that

$$\hat{\gamma} - \gamma = \frac{1}{n}\sum_{i=1}^{n} f_{ic} + o_p(n^{-1/2}) \tag{A.3}$$

and

$$\widehat{\Lambda}_{0c}(t) - \Lambda_{0c}(t) = \frac{1}{n}\Lambda_{0c}(t)\sum_{i=1}^{n} b_{ic}(t) + o_p(n^{-1/2}), \quad t \le \tau. \tag{A.4}$$

Note that the estimator $\hat{\gamma}$ is no longer efficient since the latent variable u_i's should be estimated and thus enlarge the variance of $\hat{\gamma}$.

The consistency of $\hat{\beta}$ follows from the two facts: (1) $\hat{U}(\beta)$ tends to $U(\beta)$ in probability as n tends to infinity; (2) the solution $\tilde{\beta}$ to $U(\beta) = 0$ is unique and consistent. Now we turn to prove the asymptotical normality of the proposed estimator $\hat{\beta}$. To this end, it suffices to prove that the estimating function (10) can be written as the summation of n independent and identically distributed zero-mean random variables divided by $\sqrt{n}$, plus some negligible errors.

For $i = 1, \ldots, n$, denote $\hat{x}_{4i} = (x_i', \hat{u}_i, v_i)'$, $m_i = K_i^*/\{\Lambda_{0h}(C_i)\exp(x_i'\alpha_1)\}$, and $x_{4i} = (x_i', \log(m_i), v_i)'$. Let $\hat{S}_{j,4}(\beta; t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\{\hat{x}_{4i}'(\beta + \hat{\alpha})\}\hat{x}_{4i}^{\otimes j}$ and $S_{j,4}(\beta; t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\exp\{x_{4i}'(\beta + \alpha)\}x_{4i}^{\otimes j}$ for $j = 0, 1$, and denote the limit of $S_{j,4}(\beta; t)$ as $s_{j,4}$. To prove the asymptotical normality of $\hat{\beta}$, rewrite the estimating function (10) as

$$\hat{U}(\beta) = \sum_{j=1}^{3} \hat{U}_j(\beta), \tag{A.5}$$

where

$$\hat{U}_1(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{x_{4i} - \frac{S_{1,4}(\beta; t)}{S_{0,4}(\beta; t)}\right\} N_i(t)d\tilde{N}_i(t),$$

$$\hat{U}_2(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[(\hat{x}_{4i} - x_{4i}) - \left\{\frac{\hat{S}_{1,4}(\beta; t)}{\hat{S}_{0,4}(\beta; t)} - \frac{S_{1,4}(\beta; t)}{S_{0,4}(\beta; t)}\right\}\right] N_i(t)d\tilde{N}_i(t),$$

$$\hat{U}_3(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[(\hat{x}_{2i} - \hat{x}_{4i}) - \left\{\frac{\hat{S}_1(\beta; t)}{\hat{S}_0(\beta; t)} - \frac{\hat{S}_{1,4}(\beta; t)}{\hat{S}_{0,4}(\beta; t)}\right\}\right] N_i(t)d\tilde{N}_i(t).$$

Since $\hat{U}(\beta)$ can be written as the summation of three functions in (A.5), we only need to show that each of the three parts can be written as the summation of n independent and identically distributed zero-mean random variables divided by $\sqrt{n}$, plus some negligible errors.

For $\hat{U}_1(\beta)$, it is clear that

$$\hat{U}_1(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau \left\{ x_{4i} - \frac{s_{1,4}(\beta; t)}{s_{0,4}(\beta; t)} \right\} N_i(t) d\tilde{N}_i(t) + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{1i}(\beta) + o_p(1).$$

To analyze $\hat{U}_2(\beta)$, note that u_i's are functions of $\Lambda_{0h}(C_i)$ and α_1. Let $u_i = u_i(\Lambda_{0h}(C_i), \alpha_1)$ and $x_{4i}(\Lambda_{0h}(C_i), \alpha_1) = x_{4i}$. For $j = 0, 1$ and $i = 1, \ldots, n$, denote $b_{i1}^{(j)}(\beta; t) = \Lambda_{0h}^{-1}(C_i) \frac{\partial}{\partial s} \left[Y_i(t) \exp\{(\beta+\alpha)' x_{4i}(s, \alpha_1)\} x_{4i}^{\otimes j}(s, \alpha_1) \right] |_{s=\Lambda_{0h}(C_i)}$, $b_{i2}^{(j)}(\beta; t) = \frac{\partial}{\partial \alpha_1} \left[Y_i(t) \exp\{x_{4i}'(\beta + \alpha)\} x_{4i}^{\otimes j} \right]$, $b_{i3} = \Lambda_{0h}^{-1}(C_i) \frac{\partial}{\partial s} u_i(s, \alpha_1)|_{s=\Lambda_{0h}(C_i)}$, and $b_{i4} = \frac{\partial}{\partial \alpha_1} \left[u_i(\Lambda_{0h}(C_i), \alpha_1) \right]$. Then one obtains that

$$\hat{S}_{j,4}(\beta; t) = S_{j,4}(\beta; t) + b_{i1}^{(j)}(\beta; t) \Lambda_{0h}(C_i) \{\hat{\Lambda}_{0h}(C_i) - \Lambda_{0h}(C_i)\} + b_{i2}^{(j)}(\beta; t)(\hat{\alpha}_1 - \alpha_1) + o_p(1),$$

which yields that

$$\frac{\hat{S}_{1,4}}{\hat{S}_{0,4}} = \frac{S_{1,4}}{S_{0,4}} + \frac{1}{S_{0,4}} \sum_{i=1}^{n} \left[\left(b_{i1}^{(1)} - \tfrac{S_{1,4}}{S_{0,4}} b_{i1}^{(0)} \right) \Lambda_{0h}(C_i) \{\hat{\Lambda}_{0h}(C_i) - \Lambda_{0h}(C_i)\} + \left(b_{i2}^{(1)} - \tfrac{S_{1,4}}{S_{0,4}} b_{i2}^{(0)} \right) (\hat{\alpha}_1 - \alpha_1) \right] + o_p(1).$$

Therefore, one obtains that

$$\hat{U}_2(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau \left[a_{i1} \Lambda_{0h}(C_i) \{\hat{\Lambda}_{0h}(C_i) - \Lambda_{0h}(C_i)\} + a_{i2}(\hat{\alpha}_1 - \alpha_1) \right] \times N_i(t) d\tilde{N}_i(t) + o_p(1),$$

where $a_{i1} = e_{p+1} b_{i3} + \frac{1}{s_{0,4}} \left(b_{i1}^{(1)} - \frac{s_{1,4}}{s_{0,4}} b_{i1}^{(0)} \right)$ and $a_{i2} = e_{p+1} b_{i4} + \frac{1}{s_{0,4}} \left(b_{i2}^{(1)} - \frac{s_{1,4}}{s_{0,4}} b_{i2}^{(0)} \right)$. This, together with (A.1) and (A.2), immediately yields that

$$\hat{U}_2(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau E_j \left[\{b_{ih}(C_j) a_{j1}(\beta; t) + f_{ih} a_{j2}(\beta; t)\} N_j(t) d\tilde{N}_j(t) \right] + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{2i}(\beta) + o_p(1).$$

Now we consider $\hat{U}_3(\beta)$ in (A.5). First, note that $\hat{x}_{2i} - \hat{x}_{4i} = e_{p+2}(\hat{v}_i - v_i)$. Define $b_{i5} = Y_i(t)\exp\{x_{4i}'(\beta+\alpha)\}\{e_{p+2} + x_{4i}(\beta+\alpha)'e_{p+2}\}$ and $b_{i6} = Y_i(t)\exp\{x_{4i}'(\beta+\alpha)\}(\beta+\alpha)'e_{p+2}$. Using the same arguments to the expressions of $\hat{U}_2(\beta)$, one obtains that

$$\begin{aligned}
\hat{U}_3(\beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\Bigg[e_{p+2}(\hat{v}_i - v_i)\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{S_{0,4}}b_{j5} - \frac{S_{1,4}}{S_{0,4}^2}b_{j6}\right)\\
&\qquad\times(\hat{v}_j - v_j)\Bigg]N_i(t)d\tilde{N}_i(t)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}(\hat{v}_i - v_i)\Bigg[e_{p+2}N_i(t)d\tilde{N}_i(t)\\
&\qquad-\left(\frac{1}{s_{0,4}}b_{i5} - \frac{s_{1,4}}{s_{0,4}^2}b_{i6}\right)E\{N_1(t)d\tilde{N}_1(t)\}\Bigg] + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\hat{v}_i - E_i(v_i|O_i,\theta)\}\Bigg[e_{p+2}N_i(t)d\tilde{N}_i(t)\\
&\qquad-\left(\frac{1}{s_{0,4}}b_{i5} - \frac{s_{1,4}}{s_{0,4}^2}b_{i6}\right)E\{N_1(t)d\tilde{N}_1(t)\}\Bigg]\\
&\qquad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{32i}(\beta) + o_p(1), \qquad (A.6)
\end{aligned}$$

where $g_{32i}(\beta) = \int_0^{\tau}\{E_i(v_i|O_i,\theta) - v_i\}\Bigg[e_{p+2}N_i(t)d\tilde{N}_i(t) - \left(\frac{1}{s_{0,4}}b_{i5} - \frac{s_{1,4}}{s_{0,4}^2}b_{i6}\right)$ $E\{N_1(t)d\tilde{N}_1(t)\}\Bigg]$.

By (6), using Taylor series expansion as well as the similar arguments to Fan and Li (2002) and $\hat{U}_2(\beta)$, there exist vector functions $b_{ik}, k = 7, 8, 9, 10$, such that the first term of (A.6) equals to

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\hat{v}_i - E_i(v_i|O_i,\theta)\}\Bigg[e_{p+2}N_i(t)d\tilde{N}_i(t)\\
&\qquad-\left(\frac{1}{s_{0,4}}b_{i5} - \frac{s_{1,4}}{s_{0,4}^2}b_{i6}\right)E\{N_1(t)d\tilde{N}_1(t)\}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \Big[b_{i7} \Lambda_{0h}^{-1}(C_i) \{ \hat{\Lambda}_{0h}(C_i) - \Lambda_{0h}(C_i) \} + b'_{i8}(\hat{\alpha}_1 - \alpha_1) + b'_{i9}(\hat{\gamma} - \gamma) \\
&\quad + o_p(1) + b_{i10} \Lambda_{0h}^{-1}(C_i) \{ \hat{\Lambda}_{0c}(C_i) - \Lambda_{0c}(C_i) \} \Big] \Big[e_{p+2} N_i(t) d\tilde{N}_i(t) \\
&\quad - \left(\frac{1}{s_{0,4}} b_{i5} - \frac{s_{1,4}}{s_{0,4}^2} b_{i6} \right) E\{ N_1(t) d\tilde{N}_1(t) \} \Big].
\end{aligned}
$$

In fact, $E_i(v_i|O_i, \theta)$ is a function of $(\Lambda_{0h}(\cdot), \alpha_1, \gamma, \Lambda_{0c}(\cdot))$. Then $b_{ik}, k = 7, 8, 9, 10$, can be viewed as the partial derivatives of the conditional expectation. To be specific,

$$
\begin{aligned}
b_{i7}(\cdot) &= \frac{\partial}{\partial s} E_i(v_i|O_i, \theta)(s, \alpha_1, \gamma, \Lambda_{0c}(\cdot))|_{s=\Lambda_{0h}(\cdot)}, \\
b_{i8} &= \frac{\partial}{\partial \alpha_1} E_i(v_i|O_i, \theta)(\Lambda_{0h}(\cdot), \alpha_1, \gamma, \Lambda_{0c}(\cdot)), \\
b_{i9} &= \frac{\partial}{\partial \gamma} E_i(v_i|O_i, \theta)(\Lambda_{0h}(\cdot), \alpha_1, \gamma, \Lambda_{0c}(\cdot)), \\
&\text{and} \\
b_{i10}(\cdot) &= \frac{\partial}{\partial s} E_i(v_i|O_i, \theta)(\Lambda_{0h}(\cdot), \alpha_1, \gamma, s)|_{s=\Lambda_{0c}(\cdot)}.
\end{aligned}
$$

Using (A.1)–(A.4), similarly to $\hat{U}_2(\beta)$, the first term of (A.6) equals to $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{31i}(\beta) + o_p(1)$, where

$$
\begin{aligned}
g_{31i}(\beta) = \int_0^{\tau} E_j \Big[&\big\{ b_{ih}(C_j) b_{j7} + b_{j8} f_{ih} + b_{j9} f_{ic} + b_{j10} b_{ic}(C_j) \big\} \\
&\times \left\{ e_{p+2} N_i(t) d\tilde{N}_i(t) - \left(\tfrac{1}{s_{0,4}} b_{i5} - \tfrac{s_{1,4}}{s_{0,4}^2} b_{i6} \right) E[N_1(t) d\tilde{N}_1(t)] \right\} \Big] \\
&+ o_p(1)
\end{aligned}
$$

and the expectation is taken with respect to (X_j, K_j^*, U_j, V_j). This yields that $\hat{U}_3(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ g_{31i}(\beta) + g_{32i}(\beta) \} + o_p(1)$. Combining all the above results, there exists a sequence of independent and identically distributed zero-mean random variables $g_1(\beta), \ldots, g_n(\beta)$, such that

$$
\hat{U}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\beta) + o_p(1),
$$

where $g_i(\beta) = g_{1i}(\beta) + g_{2i}(\beta) + g_{31i}(\beta) + g_{32i}(\beta), i = 1, \ldots, n$. By the standard procedure, one can obtain that the unique solution $\hat{\beta}$ to $\hat{U}(\beta) = 0$ satisfies the asymptotic normality. Specifically, $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a normal random variable with mean zero and the covariance matrix $\phi^{-1}\Sigma(\phi^{-1})'$, where $\phi = -E\{\partial g_i(\beta)/\partial\beta\}$ and $\Sigma = \text{Cov}\{g_i(\beta)\}$. This completes the proof.

References

Andrews DF, Herzberg AM (1985) Data: a collection of problems from many fields for the student and research worker. Springer-Verlag, New York

Byar DP (1980) The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In: Pavane-Macaluso M, Smith PH, Edsmyr F (eds) Bladder tumors and other topics in urological oncology. Plenum, New York, pp 363–370

Cheng SC, Wei LJ (2000) Inferences for a semiparametric model with panel data. Biometrika 87:89–97

De Gruttola V, Tu X (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics 50:1003–1014

Fan J, Li R (2002) Variable selection for Cox's proportional hazards model and frailty model. Ann Stat 30:74–99

He X, Tong X, Sun J, Cook R (2008) Regression analysis of multivariate panel count data. Biostatistics 9:234–248

Huang X, Liu L (2007) A joint frailty model for survival time and gap times between recurrent events. Biometrics 63:389–397

Huang CY, Wang MC (2004) Joint modeling and estimation for recurrent event processes and failure time data. J Am Stat Assoc 99:1153–1165

Huang CY, Wang MC, Zhang Y (2006) Analysing panel count data with informative observation times. Biometrika 93:763–775

Kalbfleisch JD, Lawless JF (1985) The analysis of panel data under a Markov assumption. J Am Stat Assoc 80:863–871

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York

Lipsitz SR, Fitzmaurice GM, Ibrahim JG, Gelber R, Lipshultz S (2002) Parameter estimation in longitudinal studies with outcome-dependent follow-up. Biometrics 58:621–630

Liu L, Huang X (2007) The use of Gaussian quadrature for estimation sin frailty proportional hazards models. Stat Med. doi:10.1002/sim.3077

Liu L, Huang X, O'Quigley J (2007) Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. Biometrics. doi: 10.1111/j.1541-0420.2007.00954.x

Liu L, Wolfe RA, Huang X (2004) Shared frailty models for recurrent events and a terminal event. Biometrics 60:747–756

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. J Roy Stat Soc Ser B 44:226–233

Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P (2007) Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. Biostatistics 8:708–721

Roy J, Lin X (2002) Analysis of multivariate longitudinal outcomes with non-ignorable dropouts and missing covariates: changes in methadone treatment practices. J Am Stati Assoc 97:40–52

Song X, Davidian M, Tsiatis AA (2002) A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. Biometrics 58:742–753

Sun J, Kalbfleisch JD (1995) Estimation of the mean function of point processes based on panel count data. Stat Sinica 5:279–290

Sun J, Tong X, He X (2007) Regression analysis of panel count data with dependent observation times. Biometrics 63:1053–1059

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

Thall PF, Lachin JM (1988) Analysis of recurrent events: Nonparametric methods for random-interval count data. J Am Stat Assoc 83:339–347

Wang MC, Qin J, Chiang CT (2001) Analyzing recurrent event data with informative censoring. J Am Stat Assoc 96:1057–1065

Wellner JA, Zhang Y (1998) Large sample theory for an estimator of the mean of a counting process with panel count data. Technical Report, Department of Statistics, University of Washington, Seattle

Wellner JA, Zhang Y (2000) Two estimators of the mean of a counting process with panel count data. Ann Stat 28:779–814

Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53:330–339

Ye Y, Kalbfleisch JD, Schaubel DE (2007) Semiparametric analysis of correlated recurrent and terminal events. Biometrics 63:78–87

Zhang Y (2002) A semiparametric pseudolikelihood estimation method for panel count data. Biometrika 89:39–48