Computational Statistics and Data Analysis 83 (2015) 151–167

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A class of transformed hazards models for recurrent gap times

Fangyuan Kang[a], Liuquan Sun[a], Xingqiu Zhao[b,*]

[a] Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China

[b] Department of Applied Mathematics, The Hong Kong Polytechnic University Shenzhen Research Institute, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 10 July 2013
Received in revised form 17 September 2014
Accepted 5 October 2014
Available online 16 October 2014

Keywords:
Estimating equations
Gap times
Model checking
Recurrent events
Transformed hazards model

ABSTRACT

In this article, a class of transformed hazards models is proposed for recurrent gap time data, including both the proportional and additive hazards models as special cases. An estimating equation-based inference procedure is developed for the model parameters, and the asymptotic properties of the resulting estimators are established. In addition, a lack-of-fit test is presented to assess the adequacy of the model. The finite sample behavior of the proposed estimators is evaluated through simulation studies, and an application to a clinic study on chronic granulomatous disease (CGD) is illustrated.

1. Introduction

Recurrent events data are commonly encountered in medical and observational studies where each subject may experience a particular event repeatedly over time. Examples of such events include repeated hospitalization, multiple infection episodes, tumor recurrences, recurrent economic recessions, and repeated breakdowns of an automobile. In these studies, it is often of interest to assess the effects of covariates on certain features of the recurrent event times. The statistical analysis of such data is challenging due to the dependence of the recurrent event times within each individual and the presence of censoring such as the loss to follow-up. To analyze recurrent event data, the focus can be laid on two types of time scale: the time since enrollment and the time between two successive recurrent events (i.e., the gap time).

When the time since enrollment is used as time index, recurrent events of a subject are modeled as the realization of an underlying counting process (Cook and Lawless, 2007), and a variety of statistical methods has been proposed in the literature. For example, Prentice et al. (1981), Andersen and Gill (1982) and Zeng and Lin (2006) proposed some intensity-based methods. Nielsen et al. (1992), Murphy (1995) and Zeng and Lin (2007) developed some frailty model approaches. Lawless and Nadeau (1995), Lin et al. (2000), Schaubel et al. (2006) and Sun et al. (2011) considered some marginal means and rates models. Cook and Lawless (2007) provided an excellent review of statistical methods for the analysis of this type of data.

In many applications, however, the gap time is a natural outcome of interest (Gail et al., 1980). Some methods have been developed for the analysis of recurrent gap time data (Huang and Chen, 2003; Schaubel and Cai, 2004; Strawderman, 2005; Luo and Huang, 2011). For example, Huang and Chen (2003), Schaubel and Cai (2004) and Darlington and Dixon (2013)

* Corresponding author.
E-mail address: xingqiu.zhao@polyu.edu.hk (X. Zhao).

http://dx.doi.org/10.1016/j.csda.2014.10.005

proposed the proportional hazards model for the gap times. Chang (2004) and Strawderman (2005) considered some accelerated gap time models. Sun et al. (2006) discussed the additive hazards model for the gap times. In addition, some authors suggested various nonparametric models for the gap time distribution (e.g., Lin et al., 1999; Wang and Chang, 1999; Peña et al., 2001 and Du, 2009). Luo and Huang (2011) demonstrated that many existing methods for recurrent gap time data can be viewed as weighted risk-set methods.

Recently, the semiparametric transformation models have been studied extensively in survival analysis (e.g., Chen et al., 2002; Zeng et al., 2005 and Zeng and Lin, 2007). However, there is a dearth of suitable transformation models for the analysis of recurrent gap time data. Lu (2005) studied the semiparametric linear transformation models for the gap times, which include the proportional hazards and proportional odds models as special case. Note that this class of models does not contain the additive hazards model as a special case. For classical survival data, Zeng et al. (2005) proposed a class of transformed hazards models which encompasses the proportional and additive hazards models and which accommodates time-varying covariates. In this paper, we consider this class of transformed hazards models for the analysis of recurrent gap time data, and propose an estimation procedure for the model parameters, which is easy to implement.

The remainder of the paper is organized as follows. In Section 2, we introduce data structure and the proposed models. Estimation procedures are presented for the model parameters, and the asymptotic properties of the proposed estimators are established. In Section 3, we develop a technique for checking the adequacy of the proposed model. Section 4 reports some results from simulation studies conducted for evaluating the proposed methods. An application to a clinic study on CGD is provided in Section 5, and some concluding remarks are given in Section 6. All proofs are given in the Appendix.

2. Model and estimation procedure

2.1. The model

Consider a longitudinal study that involves n independent subjects, each of which experiences recurrences of the same event (Huang and Chen, 2003; Luo and Huang, 2011). For subject i, let T_{ij} denote the time from the $(j-1)$th to the jth occurrence of the event. That is, $T_{i1} + \cdots + T_{ij}$ is the jth recurrent event time. Also let Z_i denote the p-dimensional vector of covariates associated with subject i, and C_i the follow-up or censoring time. Let $N_i = \{T_{ij} : j = 1, 2, \ldots\}$. Assume that $\{N_i, C_i, Z_i\}$ $(i = 1, \ldots, n)$ are independent and identically distributed (i.i.d.), and N_i is independent of C_i given Z_i. Define M_i to be the index of observed gap times for subject i, which satisfies

$$\sum_{j=1}^{M_i-1} T_{ij} \le C_i \quad \text{and} \quad \sum_{j=1}^{M_i} T_{ij} > C_i,$$

where $\sum_{j=1}^{0} \cdot \equiv 0$. Then observed data are $\{T_{i1}, \ldots, T_{i,M_i-1}, C_i, Z_i\}$. That is, the first $M_i - 1$ gap times are observed, but T_{i,M_i} is censored at $T^{+}_{i,M_i} = C_i - \sum_{j=1}^{M_i-1} T_{ij}$.

Following Huang and Chen (2003), we assume that each individual recurrent event process is a renewal process, which implies that for a given i, $\{T_{ij},\ j = 1, 2, \ldots\}$ are i.i.d., and that for given $(C_i, M_i, T^{+}_{i,M_i})$, the observed complete gap times $\{T_{ij},\ j = 1, \ldots, M_i - 1\}$ are identically distributed (Wang and Chang, 1999).

Let $\lambda_{ij}(t|Z_i)$ be the hazard function of T_{ij} given Z_i. The proposed transformed hazards models take the form

$$\lambda_{ij}(t|Z_i) = H\{\lambda_0(t) + \beta_0' Z_i\}, \tag{1}$$

where $\lambda_0(t)$ is an unknown function, β_0 is a $p \times 1$ vector of unknown regression parameters, and $H(\cdot)$ is pre-specified and assumed to be twice continuously differentiable and strictly increasing. Model (1) defines a very rich family of models through the link function $H(\cdot)$, which includes the proportional hazards model ($H(x) = \exp(x)$) and the additive hazards models ($H(x) = x$). One example of $H(\cdot)$ is the Box–Cox transformation, in which $H(\cdot)$ is given by $H(x) = \{(1 + x)^s - 1\}/s$ for $s \ge 0$ with $s = 0$ corresponding to $H(x) = \log(x + 1)$. Another useful class is the logarithmic transformations, which are given by $H(x) = \log(1 + \gamma x)/\gamma$ for $\gamma \ge 0$ with $\gamma = 0$ corresponding to $H(x) = x$.

2.2. Inference procedure

Our inference procedure is based on the establishment of a connection between a subset of the observed gap times and clustered survival data. Let $\Delta_i = I(M_i > 1)$, $S_i = \max(M_i - 1, 1)$, and

$$X_{ij} = \begin{cases} T_{ij} & \text{if } \Delta_i = 1, \\ T^{+}_{ij} & \text{if } \Delta_i = 0, \end{cases} \quad j = 1, \ldots, S_i.$$

Then $\{X_{ij}, \Delta_i, Z_i,\ j = 1, \ldots, S_i\}$ $(i = 1, \ldots, n)$ can be treated as clustered survival data. Since the cluster size is informative, the censored gap time needs to be removed for $M_i > 1$ (Wang and Chang, 1999; Huang and Chen, 2003).

Define $N_{ij}(t) = \Delta_i I(X_{ij} \le t),\ Y_{ij} = I(X_{ij} \ge t)$, and

$$M_{ij}(t; \beta, \lambda) = N_{ij}(t) - \int_0^t Y_{ij}(u) H\{\lambda(u) + \beta' Z_i\} du.$$

Then under model (1), $M_{ij}(t;\beta_0,\lambda_0)$ ($j=1,\dots,S_i;\ i=1,\dots,n$) are zero-mean stochastic processes but not martingales, since the hazard function is not modeled conditionally on the recurrent event process. Based on this fact and using the generalized estimating equation approach (Liang and Zeger, 1986), we propose the following estimating functions for $\lambda_0(t)$ and β_0:

$$\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}\int_0^t\Big[dN_{ij}(u)-Y_{ij}(u)H\{\lambda(u)+\beta' Z_i\}du\Big]=0, \tag{2}$$

and

$$\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}\int_0^\tau Z_i\Big[dN_{ij}(u)-Y_{ij}(u)H\{\lambda(u)+\beta' Z_i\}du\Big]=0, \tag{3}$$

where τ is a pre-specified constant such that $P(C_i\ge\tau)>0$.

Because the first term on the left-hand side of (2) represents a pure jump process while the second is absolutely continuous, this equation has no solution. However, the above estimating equations will be the starting point for our estimation. To this end, we proceed by a Taylor expansion of $H\{\lambda(t)+\beta' Z_i\}$ around the current value of estimates $\lambda^{(k)}(t)$ and $\beta^{(k)}$ to get the approximated estimating equations:

$$\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}\int_0^t\Big[dN_{ij}(u)-Y_{ij}(u)H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}du-Y_{ij}(u)\dot H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}\{\lambda(u)-\lambda^{(k)}(u)\}du$$
$$-Y_{ij}(u)\dot H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}Z_i'(\beta-\beta^{(k)})du\Big]=0, \tag{4}$$

and

$$\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}\int_0^\tau Z_i\Big[dN_{ij}(u)-Y_{ij}(u)H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}du-Y_{ij}(u)\dot H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}\{\lambda(u)-\lambda^{(k)}(u)\}du$$
$$-Y_{ij}(u)\dot H\{\lambda^{(k)}(u)+Z_i'\beta^{(k)}\}Z_i'(\beta-\beta^{(k)})du\Big]=0, \tag{5}$$

where $\dot H(x)=dH(x)/dx$. Define

$$dM_i(t;\beta,\lambda)=\frac{1}{S_i}\sum_{j=1}^{S_i}dM_{ij}(t;\beta,\lambda),$$

$$\phi_i(t;\beta,\lambda)=\frac{1}{S_i}\sum_{j=1}^{S_i}Y_{ij}(t)\dot H\{\lambda(t)+Z_i'\beta\},$$

$$S_0(t;\beta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}Y_{ij}(t)\dot H\{\lambda(t)+Z_i'\beta\},$$

$$S_z(t;\beta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{S_i}\sum_{j=1}^{S_i}Y_{ij}(t)\dot H\{\lambda(t)+Z_i'\beta\}Z_i,$$

$$E_{zz}(t;\beta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\{Z_i-E_z(t;\beta,\lambda)\}\phi_i(t;\beta,\lambda)\{Z_i-E_z(t;\beta,\lambda)\}',$$

and $dM(t;\beta,\lambda)=(dM_1(t;\beta,\lambda),\dots,dM_n(t;\beta,\lambda))'$, where $E_z(t;\beta,\lambda)=S_z(t;\beta,\lambda)/S_0(t;\beta,\lambda)$. Also let $\mathbf{Z}$ be a $n\times p$ matrix with rows Z_i', and $\bar{\mathbf{Z}}(t;\lambda,\beta)$ be a $n\times p$ matrix with rows $E_z(t;\lambda,\beta)'$. Solving (4) for $\lambda(t)$ and inserting it into (5), we get the $(k+1)$th iterative estimator for β_0: $\beta^{(k+1)}=Q(\beta^{(k)})$, where

$$Q(\beta^{(k)})=\beta^{(k)}+\frac{1}{n}A(\tau;\beta^{(k)},\lambda^{(k)})^{-1}\int_0^\tau\{\mathbf{Z}-\bar{\mathbf{Z}}(t;\beta^{(k)},\lambda^{(k)})\}'dM(t;\beta^{(k)},\lambda^{(k)}), \tag{6}$$

and $A(\tau;\beta,\lambda)=\int_0^\tau E_{zz}(t;\beta,\lambda)dt$.

For $\lambda_0(t)$, it is not a good idea to try to iterate towards a solution for each time point t, because the information about any particular time point is limited. So, a smoothing solution is needed. In the following, we focus on $\Lambda_0(t)=\int_0^t\lambda_0(u)du$ (e.g., Martinussen et al., 2002; Zeng and Lin, 2006). Using the updated version $\beta^{(k+1)}$ and solving (4), we obtain the $(k+1)$th iterative estimator for $\Lambda_0(t)$: $\Lambda^{(k+1)}(t) = \Psi(\Lambda^{(k)})(t)$, where

$$\Psi(\Lambda^{(k)})(t) = \int_0^t \lambda^{(k)}(u)du + \frac{1}{n}\sum_{i=1}^n \int_0^t S_0^{-1}(u;\beta^{(k)},\lambda^{(k)})dM_i(u;\beta^{(k)},\lambda^{(k)}) - \int_0^t E_z(u;\beta^{(k)},\lambda^{(k)})'du(\beta^{(k+1)} - \beta^{(k)}). \tag{7}$$

Here for simplicity, $\lambda^{(k)}(t)$ is taken to be a simple kernel estimator based on $\Lambda^{(k)}(t)$ with positive bandwidth h, that is,

$$\lambda^{(k)}(t) = \int \frac{1}{h}K\Big(\frac{u-t}{h}\Big)d\Lambda^{(k)}(u),$$

and K is a symmetric kernel function with a compact support.

Given the initial estimators $\lambda^{(k)}(t)$ and $\beta^{(k)}$, the estimation procedure can be summarized as follows.

Step 1. Use Eq. (6) to obtain $\beta^{(k+1)}$.

Step 2. Use $\beta^{(k+1)}$ and Eq. (7) to obtain $\Lambda^{(k+1)}$.

Step 3. Smooth $\Lambda^{(k+1)}$ to obtain $\lambda^{(k+1)}$, and return to Step 1 with updated estimators until convergence.

2.3. Asymptotic properties

We now describe the asymptotic behavior of the proposed estimators, and the results are summarized in the following theorem with the proof given in Appendix A.

Theorem 1. *Under the regularity conditions* (C1)–(C5) *stated in Appendix A, we have*

(i) *with probability tending to one,* (6) *and* (7) *have solutions* $Q(\hat\beta) = \hat\beta$ *and* $\Psi(\hat\Lambda) = \hat\Lambda$ *such that* $\|\hat\beta - \beta_0\| = O_p(n^{-1/2})$ *and* $\sup_{0\le t\le\tau}|\hat\Lambda(t) - \Lambda_0(t)| = O_p(n^{-1/2})$.

(ii) $n^{1/2}(\hat\beta - \beta_0)$ *is asymptotically normal with zero mean and a covariance matrix that can be consistently estimated by* $\hat A^{-1}\hat\Sigma\hat A^{-1}$, *where* $\hat\Sigma = n^{-1}\sum_{i=1}^n \hat\xi_i\hat\xi_i'$,

$$\hat\xi_i = \int_0^\tau \{Z_i - E_z(t;\hat\beta,\hat\lambda)\}dM_i(t;\hat\beta,\hat\lambda),$$

$$\hat A = \int_0^\tau E_{zz}(t;\hat\beta,\hat\lambda)dt,$$

and

$$\hat\lambda(t) = \int \frac{1}{h}K\Big(\frac{u-t}{h}\Big)d\hat\Lambda(u).$$

(iii) $n^{1/2}\{\hat\Lambda(t) - \Lambda_0(t)\}$ $(0 \le t \le \tau)$ *converges weakly to a zero-mean Gaussian process whose covariance function at* (s,t) *can be consistently estimated by*

$$\hat\Gamma(s,t) = \frac{1}{n}\sum_{i=1}^n \hat\eta_i(s)\hat\eta_i(t),$$

where

$$\hat\eta_i(t) = \int_0^t S_0^{-1}(u;\hat\beta,\hat\lambda)dM_i(u;\hat\beta,\hat\lambda) - \int_0^t E_z(u;\hat\beta,\hat\lambda)'\hat A^{-1}\hat\xi_i du.$$

3. Goodness-of-fit tests

First we consider checking a possible misspecification of the functional form of covariates. Let Z_{ji} be the jth component of Z_i. We define the cumulative sum of residuals

$$\mathcal{F}_j(\tau,z) = n^{-1/2}\sum_{i=1}^n \int_0^\tau I\{Z_{ji} \le z\}dM_i(u;\hat\beta,\hat\lambda)$$

in the same way as used in Lin et al. (1993). When K is an rth-order $(r > 3)$ kernel function, which satisfies $\int K(u)du = 1$, $\int u^m K(u)du = 0,\ m = 1,\dots,r-1$, and $\int u^r K(u)du \ne 0$, we show in Appendix B that the null distribution of $\mathcal{F}_j(\tau,z)$ can be approximated by the zero-mean Gaussian process

$$\tilde{\mathcal{F}}_j(\tau,z) = n^{-1/2}\sum_{i=1}^n \Big[\int_0^\tau I\{Z_{ji}\le z\}dM_i(u;\hat\beta,\hat\lambda) - \int_0^\tau \hat\Phi_j(u,z)d\hat\eta_i(u) - \hat B_j(\tau,z)'\hat A^{-1}\hat\xi_i\Big],$$

where

$$\hat{\Phi}_j(u,z) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_{ji} \le z\}\phi_i(u;\hat{\beta},\hat{\lambda}),$$

and

$$\hat{B}_j(\tau,z) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} I\{Z_{ji} \le z\}\phi_i(u;\hat{\beta},\hat{\lambda})Z_i du.$$

In Section 5, to check the functional form for the jth component of Z_i, we simply plot the residual $M_i(\tau;\hat{\beta},\hat{\lambda})$ against Z_{ji}.

Furthermore, we propose a lack-of-fit test for assessing the adequacy of model (1). Following Lin et al. (2000), we consider the following cumulative sum of residuals:

$$\mathcal{F}(t,z) = n^{-1/2}\sum_{i=1}^{n}\int_0^{t} I\{Z_i \le z\}dM_i(u;\hat{\beta},\hat{\lambda}), \tag{8}$$

where $I\{Z_i \le z\}$ means that each component of Z_i is not larger than the corresponding component of z.

The null distribution of $\mathcal{F}(t,z)$ can be approximated by the zero-mean Gaussian process

$$\tilde{\mathcal{F}}(t,z) = n^{-1/2}\sum_{i=1}^{n}\Big[\int_0^{t} I\{Z_i \le z\}dM_i(u;\hat{\beta},\hat{\lambda}) - \int_0^{t}\hat{\Phi}(u,z)d\hat{\eta}_i(u) - \hat{B}(t,z)'\hat{A}^{-1}\hat{\xi}_i\Big], \tag{9}$$

where

$$\hat{\Phi}(u,z) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i \le z\}\phi_i(u;\hat{\beta},\hat{\lambda}),$$

and

$$\hat{B}(t,z) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{t} I\{Z_i \le z\}\phi_i(u;\hat{\beta},\hat{\lambda})Z_i du.$$

Note that it is impossible to evaluate the above distribution analytically because the limiting process of $\mathcal{F}(t,z)$ does not have independent increments. To overcome this difficulty, we propose to use the following resampling approach (e.g., Lin et al., 2000). Let $(G_1,\ldots,G_n)$ be independent standard normal variables independent of the data. Then it can be shown that the distribution of $\mathcal{F}(t,z)$ can be approximated by that of the zero-mean Gaussian process

$$\hat{\mathcal{F}}(t,z) = n^{-1/2}\sum_{i=1}^{n}\Big[\int_0^{t} I\{Z_i \le z\}dM_i(u;\hat{\beta},\hat{\lambda}) - \int_0^{t}\hat{\Phi}(u,z)d\hat{\eta}_i(u) - \hat{B}(t,z)'\hat{A}^{-1}\hat{\xi}_i\Big]G_i.$$

Thus, we can obtain a large number of realizations from $\hat{\mathcal{F}}(t,z)$ by repeatedly generating the standard normal random sample $(G_1,\ldots,G_n)$ while fixing the observed data, and use the empirical distribution of these realizations to approximate the distribution of $\mathcal{F}(t,z)$. To assess the fit of model (1), we can apply the supremum test statistic $U = \sup_{0\le t\le\tau,z}|\mathcal{F}(t,z)|$, whose p-value can be obtained by comparing the observed value of $\sup_{0\le t\le\tau,z}|\mathcal{F}(t,z)|$ to a large number of realizations from $\sup_{0\le t\le\tau,z}|\hat{\mathcal{F}}(t,z)|$. Similarly, we can obtain p-values of $U_j = \sup_{0\le t\le\tau,z}|\mathcal{F}_j(t,z)|$ for $j = 1,\ldots,p$. By applying the proposed test statistics to a given set of data, the model with the largest p-value fits the data best among all candidate models.

4. Simulation studies

Simulation studies were conducted to examine the finite sample properties of the proposed estimators. In the study, a heterogeneous mixture of individual renewal processes was used with model (1). Specifically, the baseline gap time T_{ij}^0 was assumed to follow the standard exponential distribution, and set to be $-\ln\{1-\Phi(A_i+B_{ij})\}$, where Φ is the cumulative distribution function of the standard normal distribution, A_i and B_{ij} are independent normal random variables with mean zeros and variances ρ and $1-\rho$, respectively, with $\rho \in [0,1]$. Here the parameter ρ dictates the heterogeneity of between-individual, and $1-\rho$ controls the heterogeneity of between-episodes within an individual. Given the baseline gap times, general gap times were taken as $T_{ij}^0/H(0.5+\beta_0 Z_i)$ with $\beta_0 = 0.5$, where Z_i is a uniform random variable on $(0,1)$.

We considered three choices for H: an identity transformation $H_1(x) = x$, an exponential transformation $H_2(x) = 0.3\exp(x)$, and a Box–Cox transformation $H_3(x) = \{(1+x)^s - 1\}/s$, with $s = 0, 0.5, 1, 2$ and 3. Note that model (1) reduces to the additive hazards model studied by Sun et al. (2006) when $H(x) = H_1(x)$ or $H(x) = H_3(x)$ with $s = 1$, and reduces to the proportional hazards model studied by Huang and Chen (2003) when $H(x) = H_2(x)$.

For each case, we considered $\rho = 0.25, 0.5$ or 0.75. The censoring time C_i was taken as the minimum of the uniform distribution on $(0,\nu)$ and τ, with $\tau = 2$ and ν varying to yield an average recurrence number of 1–3 for different model

Table 1
Simulation results for the estimation of β_0 under $H_1(x)$.

n	ρ	Proposed method				SPS method			
		Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
100	0.25	0.040	0.371	0.355	0.942	0.040	0.371	0.356	0.940
	0.5	0.025	0.366	0.354	0.942	0.025	0.366	0.356	0.943
	0.75	0.002	0.357	0.353	0.947	0.002	0.357	0.354	0.944
200	0.25	0.022	0.256	0.244	0.945	0.022	0.256	0.245	0.945
	0.5	0.012	0.253	0.245	0.941	0.012	0.253	0.246	0.940
	0.75	0.003	0.246	0.243	0.945	0.003	0.246	0.244	0.948

Note: "SPS method" stands for the method of Sun et al. (2006).

Table 2
Simulation results for the estimation of β_0 under $H_2(x)$.

n	ρ	Proposed method				HC method			
		Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
100	0.25	0.014	0.548	0.557	0.962	0.014	0.548	0.533	0.948
	0.5	−0.033	0.545	0.557	0.953	−0.033	0.545	0.535	0.945
	0.75	−0.003	0.554	0.561	0.956	−0.003	0.554	0.542	0.939
200	0.25	0.010	0.378	0.384	0.951	0.010	0.378	0.373	0.945
	0.5	0.012	0.380	0.384	0.958	0.012	0.380	0.374	0.957
	0.75	−0.011	0.388	0.388	0.955	−0.011	0.388	0.380	0.951

Note: "HC method" stands for the method of Huang and Chen (2003).

Table 3
Simulation results for the estimation of β_0 under $H_3(x)$.

s	ρ	$n = 100$				$n = 200$			
		Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
0	0.25	0.035	0.454	0.449	0.947	0.012	0.310	0.308	0.950
	0.5	−0.008	0.449	0.449	0.958	0.010	0.298	0.309	0.955
	0.75	−0.010	0.425	0.440	0.966	−0.021	0.303	0.303	0.952
0.5	0.25	0.024	0.437	0.454	0.955	0.012	0.298	0.295	0.945
	0.5	0.021	0.439	0.455	0.958	0.008	0.296	0.295	0.946
	0.75	0.038	0.437	0.447	0.951	−0.026	0.284	0.290	0.951
1	0.25	0.033	0.380	0.366	0.940	0.017	0.255	0.246	0.951
	0.5	0.020	0.374	0.365	0.948	0.013	0.252	0.245	0.942
	0.75	0.005	0.378	0.364	0.938	−0.008	0.250	0.245	0.942
2	0.25	0.045	0.291	0.323	0.951	0.020	0.190	0.197	0.953
	0.5	0.034	0.292	0.326	0.951	0.020	0.206	0.198	0.932
	0.75	0.031	0.302	0.331	0.943	0.000	0.203	0.196	0.945
3	0.25	0.026	0.261	0.280	0.946	0.009	0.179	0.177	0.946
	0.5	0.033	0.260	0.279	0.940	0.017	0.175	0.176	0.953
	0.75	0.015	0.260	0.283	0.940	0.005	0.182	0.177	0.941

parameters. For the estimation of $\lambda_0(t)$, we used the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ with bandwidth $h = 0.05$ for all simulations. Note that the commonly used cross-validation criterion cannot be directly used for the optimal bandwidth selection, since it is difficult to derive the mean square error of the kernel estimator $\hat{\lambda}(t)$ for recurrent gap time data. It deserves a further study to develop an optimal bandwidth selection method for this case. Here, we selected the bandwidth based on the average distance between data points such that several data points lay within one bandwidth at a given data point. In our simulation with $n = 100$, we find that there are 1.5, 4 and 7 data points on average within one bandwidth at a given data point for $h = 0.01$, 0.03 and 0.05, respectively. We also conducted simulations for $h = 0.01$ and 0.03. When $h = 0.01$, the estimated variance is too large; while $h = 0.03$, the obtained simulation results are similar to those with $h = 0.05$. Therefore, the bandwidth can be chosen to be 0.03 or 0.05 in the settings considered here. For comparison, we also considered the method of Sun et al. (2006) with $H_1(x)$, and the method of Huang and Chen (2003) with $H_2(x)$ under the same setup as above. The results presented below are based on 1000 replications with sample sizes $n = 100$ and $n = 200$, and the final estimates were reached at convergence.

Tables 1–3 present the simulation results for the estimate of β_0 under $H_1(x)$, $H_2(x)$ and $H_3(x)$, respectively. In these tables, Bias, SSE, ESE, and CP stand for the sample mean of the estimate minus the true value, the sampling standard error of the estimate, the sampling mean of the estimated standard error, and the 95% empirical coverage probability for β_0 based on a normal approximation, respectively. It can be seen from Tables 1–3 that the proposed estimation procedures perform well for the situations considered here. It appears that the proposed estimators are practically unbiased, and there is a good agreement between the estimated and empirical standard errors. The coverage probabilities seem reasonable, and the results become better when the sample size increases from 100 to 200.

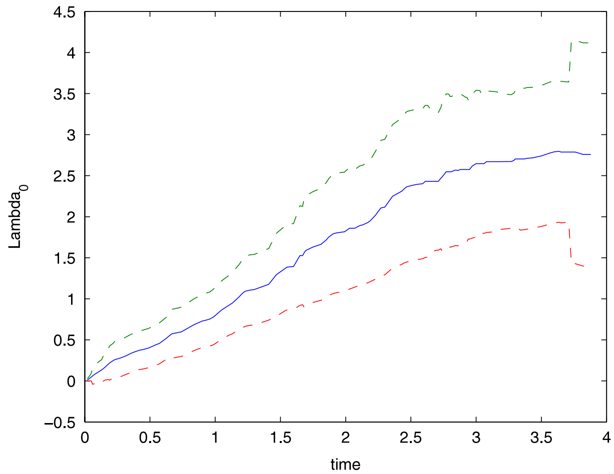

Fig. 1. Estimates are $\Lambda_0(t)$ and their pointwise 95% confidence bands for the $s = 0$ model considered.

Note that under $H_1(x)$, our estimator has the same expression as that of Sun et al. (2006). Thus, the two methods provide the same estimates and the same sampling standard errors of the estimates. Also the variances are almost the same for the two methods. Under $H_2(x)$, our method and Huang and Chen (2003)'s method give reasonable and comparable estimates, and the variances of our method are only slightly larger than those of Huang and Chen (2003)'s method.

We also considered other setups and the results were similar to those given above. In particular, we obtained the similar simulation results for different values of bandwidth h, which seems to indicate that the proposed estimation method is robust with respect to the bandwidth choice.

5. An application

In this section, we apply the proposed method to the multiple-infection data taken from the CGD study. The data are given in Appendix D of Fleming and Harrington (1991) and were analyzed by Lin et al. (2000) and Sun et al. (2011) among others. The CGD is a group of inherited rare disorders of the immune function characterized by recurrent pyogenic infections which usually occur early in life and may lead to death in childhood. In order to investigate the ability of gamma interferon to reduce the hazard of serious infections requiring hospitalization, a double-blinded clinical trial was conducted in which patients were randomized to either placebo or gamma interferon group, and a total of 128 patients were enrolled into the study. The data set includes the dates of randomization and each serious infection during the follow-up for each patient. By the end of the study, 30 of the 65 patients in the placebo group and 14 of 63 in the gamma interferon group had experienced at least one serious infection.

For the analysis, we defined Z_{1i} as the treatment indicator, which took the value 1 if the subject received gamma interferon or 0 if the subject was in the placebo group, Z_{2i} as the patients age at enrollment, Z_{3i} as an indicator for use of prophylactic antibiotics at study entry (yes $= 1$, no $= 0$), and Z_{4i} as a binary indicator of gender (male $= 0$, female $= 1$). Let $Z_i = (Z_{1i}, Z_{2i}, Z_{3i}, Z_{4i})'$. For the illustration purpose, we assumed that the data can be described by model (1) with a Box–Cox transformation $H(x) = \{(1+x)^s - 1\}/s$ for $s = 0, 0.5, 1$ and 2, and an exponential transformation $H(x) = 0.3\exp(x)$. When $s = 0$, $\log(1+\lambda_0(t))$ is the baseline hazard function, and β_0 denotes the log linear influence of the covariates. When $s = 0.5$, $2((1+\lambda_0(t))^{1/2} - 1)$ is the baseline hazard function, and β_0 is the square root linear influence of the covariates. When $s = 1$, $\lambda_0(t)$ is the baseline hazard function, and β_0 denotes the linear influence of the covariates. For the estimation of $\lambda_0(t)$, we used the bivariate Gaussian-based kernel function of order 4 (Wand and Schucany, 1990): $K(u) = (3-u^2)\exp(-u^2)/(8\pi)^{1/2}$ with bandwidth $h = 1$. Using the same principle adopted in the simulation, we selected $h = 1$ such that there are 3–7 data points within one bandwidth at each given data point. Let τ be the largest observed infection time. The analysis results are summarized in Table 4. All results suggest that treatment, age and gender all have significant effects on the hazard of infections. In particular, the gamma interferon is effective in reducing the hazard of infections, the infection hazard is lower for older patients, and females are more easily infected than males. These findings are similar to those obtained by Lin et al. (2000) and Sun et al. (2011). Figs. 1–5 display the confidence interval estimates of $\Lambda_0(t)$ together with their pointwise 95% confidence bands for all models considered. For comparison, we also provided the predicted survival functions for the additive hazards model and the Cox model with the corresponding stratified Kaplan–Meier estimates in Figs. 6–7, in which the curves are stratified by treatment and gender, where the age value is taken as its median and the indicator for use of prophylactic antibiotics is taken as 0. These plots indicate that the additive hazards model provides a good fit to the data.

In order to examine which model fits the data best, we used the model checking techniques presented in Section 3 to compare the performances of the five models considered here. The estimated p-values of the proposed test statistics for these models are given in Table 4. The last column of Table 4 suggests that all five models are reasonable to fit the data, and the additive hazards model ($s = 1$) and the Cox model fit the data more adequately.

Table 4
Analysis results for CGD data.

Model	Covariate	Est	95% confidence interval	SE	p-value (U_j)	p-value (U)
$s = 0$	Treat	−0.2264	(−0.2763, −0.1765)	0.0255	0.002	
	Age	−0.0148	(−0.0151, −0.0145)	0.0001	0.002	
	Prophy	−0.1374	(−0.2875, 0.0127)	0.0766	0.003	
	Gender	0.3862	(0.2787, 0.4937)	0.0549	0.003	0.004
$s = 0.5$	Treat	−0.2132	(−0.2587, −0.1676)	0.0233	0.440	
	Age	−0.0160	(−0.0162, −0.0158)	0.0001	0.287	
	Prophy	−0.1304	(−0.2812, 0.0205)	0.0770	0.467	
	Gender	0.3563	(0.2729, 0.4397)	0.0425	0.464	0.390
$s = 1$	Treat	−0.2009	(−0.2427, −0.1591)	0.0213	0.995	
	Age	−0.0169	(−0.0171, −0.0167)	0.0001	0.985	
	Prophy	−0.1240	(−0.2670, 0.0190)	0.0730	1.000	
	Gender	0.3285	(0.2757, 0.3814)	0.0270	0.994	0.951
$s = 2$	Treat	−0.1788	(−0.2085, −0.1491)	0.0151	0.166	
	Age	−0.0179	(−0.0180, −0.0178)	0.0001	0.187	
	Prophy	−0.1127	(−0.2028, −0.0227)	0.0459	0.171	
	Gender	0.2788	(0.2234, 0.3341)	0.0282	0.178	0.263
Cox	Treat	−0.5785	(−0.8570, −0.3001)	0.1421	0.602	
	Age	−0.0554	(−0.0562, −0.0546)	0.0004	0.770	
	Prophy	−0.3598	(−0.9938, 0.2743)	0.3235	0.611	
	Gender	0.9171	(0.5375, 1.2968)	0.1937	0.604	0.901

Est: the parameter estimate; SE: the standard error estimate.

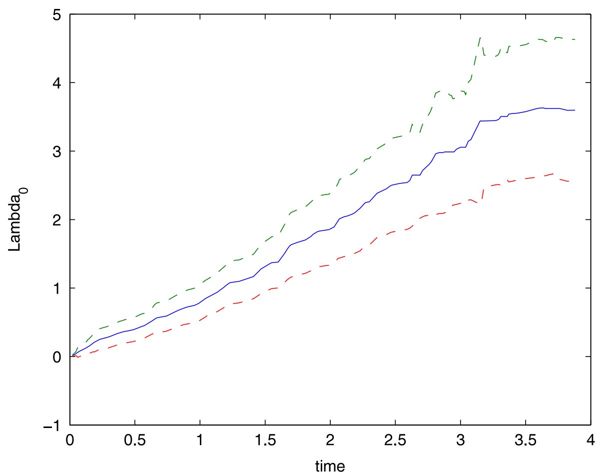

Fig. 2. Estimates are $\Lambda_0(t)$ and their pointwise 95% confidence bands for the $s = 0.5$ model considered.

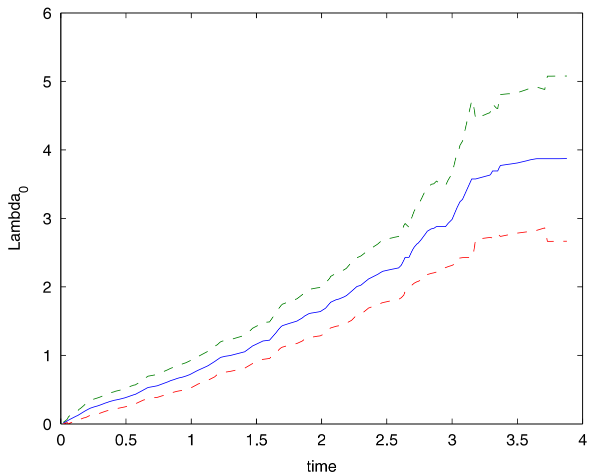

Fig. 3. Estimates are $\Lambda_0(t)$ and their pointwise 95% confidence bands for the $s = 1$ model considered.

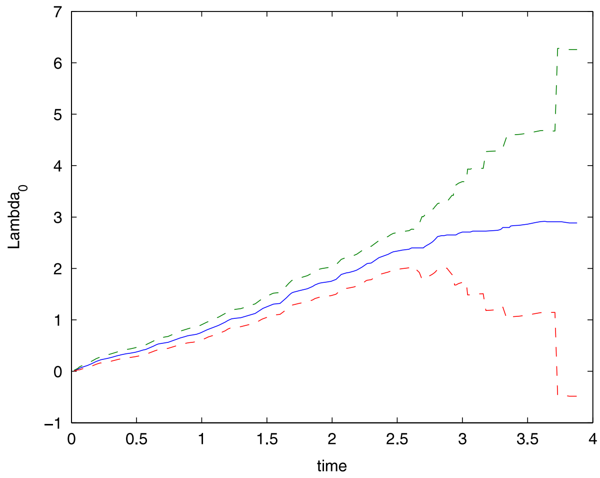

Fig. 4. Estimates are $\Lambda_0(t)$ and their pointwise 95% confidence bands for the $s = 2$ model considered.

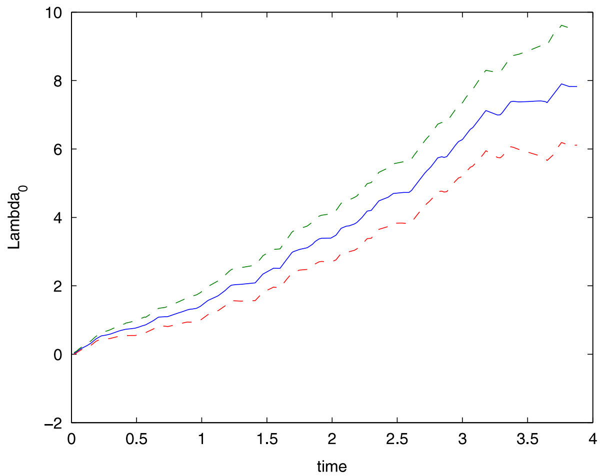

Fig. 5. Estimates are $\Lambda_0(t)$ and their pointwise 95% confidence bands for the Cox model considered.

Furthermore, for the additive hazards model and the Cox model, we checked the misspecification of the functional form of covariates. We plotted the residual $M_i(\tau; \hat{\beta}, \hat{\lambda})$ against each component of Z_i shown in Figs. 8–9. These results indicate that the residuals fluctuate around zero and seem to be random for the additive hazards model, while some residuals do not fluctuate around zero for the Cox model. Therefore, the additive hazards model may be the best one among the five models for fitting the CGD data.

6. Concluding remarks

In this article, we have proposed a class of transformed hazards models for recurrent gap time data. The new model offers great flexibility in formulating the effects of covariates on the hazards function for the gap times. An estimation procedure was proposed for the model parameters, and the asymptotic properties of the proposed estimators were derived. Simulation studies showed that the proposed method performs well for practical situations, and an illustrative example was provided.

Note that model (1) assumes a common baseline hazard function $\lambda_0(t)$ for the recurrent gap times of each subject. However, the proposed estimation procedure can be extended in a straightforward manner to deal with gap-time-specific baseline hazard functions. In this case, model (1) would become

$$\lambda_{ij}(t|Z_{ij}^*) = H\{\lambda_{0j}(t) + \beta_0' Z_{ij}^*\},$$

where $\lambda_{0j}(t)'s$ are unspecified functions, and Z_{ij}^* is a covariate vector associated with the gap times T_{ij} $i = 1, \ldots, n$ and $j \geq 1$.

Our proposed method assumes that the censoring time and the gap times are independent conditional on covariates. However, in many applications, this noninformative censoring assumption might not hold, especially when censoring could be caused by a terminal event such as death. One possible way to adjust the method for such dependent censoring is to

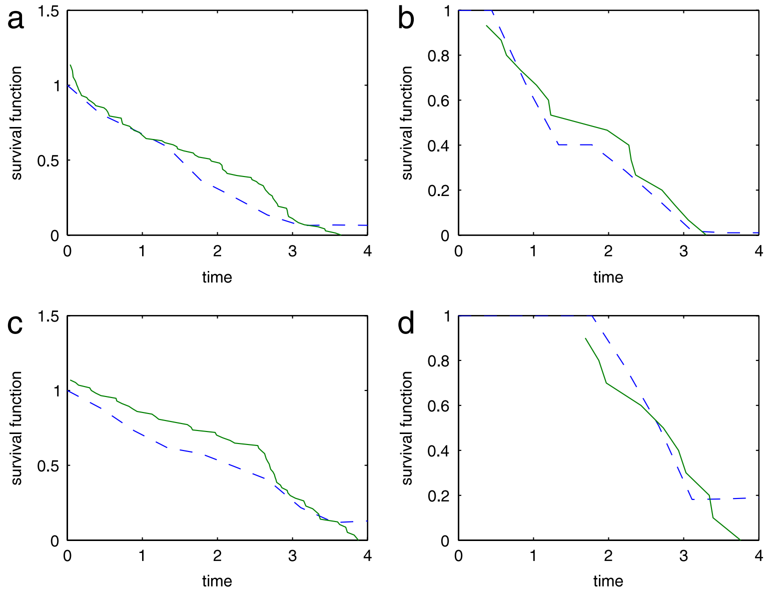

Fig. 6. $s = 1$. *Note*: Estimated survival functions (- - -) versus Kaplan–Meier estimates (—). (a) placebo, male; (b) placebo, female; (c) gamma interferon, male; (d) gamma interferon, female.

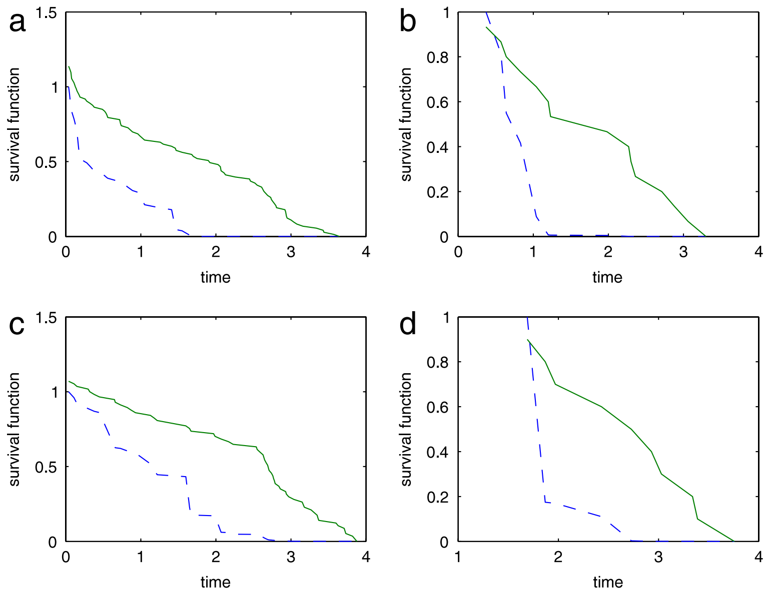

Fig. 7. Cox. *Note*: Estimated survival functions (- - -) versus Kaplan–Meier estimates (—). (a) placebo, male; (b) placebo, female; (c) gamma interferon, male; (d) gamma interferon, female.

use a joint frailty model to simultaneously analyze the gap times and the censoring time as in Huang and Liu (2007). When time-varying covariates are associated with gap times, model (1) needs to be modified as discussed in Chapter 4 of Cook and Lawless (2007). It would be interesting to extend the procedures for the class of transformed hazards models to handle these problems.

Since estimating functions (2) and (3) were given in a somewhat ad hoc fashion using the generalized estimating equation approach, it would be worthwhile to further investigate the efficiencies of the proposed estimators. If the recurrent event process is a Poisson process, then it might be possible to estimate β_0 and $\lambda_0(t)$ more efficiently by the nonparametric maximum likelihood approach, and the resulting inference procedure would be much more complicated.

Acknowledgments

The authors would like to thank the Editor, Professor Jae Chang Lee, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. Sun's research was partly supported

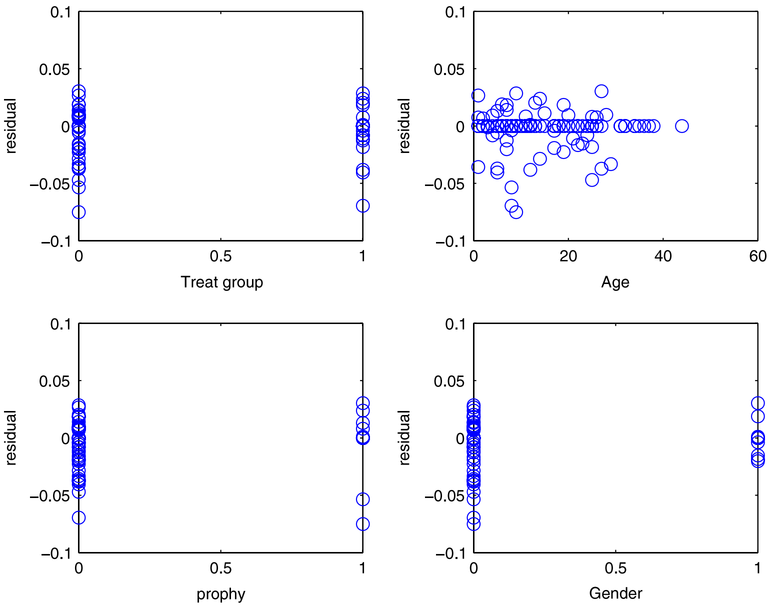

Fig. 8. Plots of residual versus covariates for the additive $s = 1$ model.

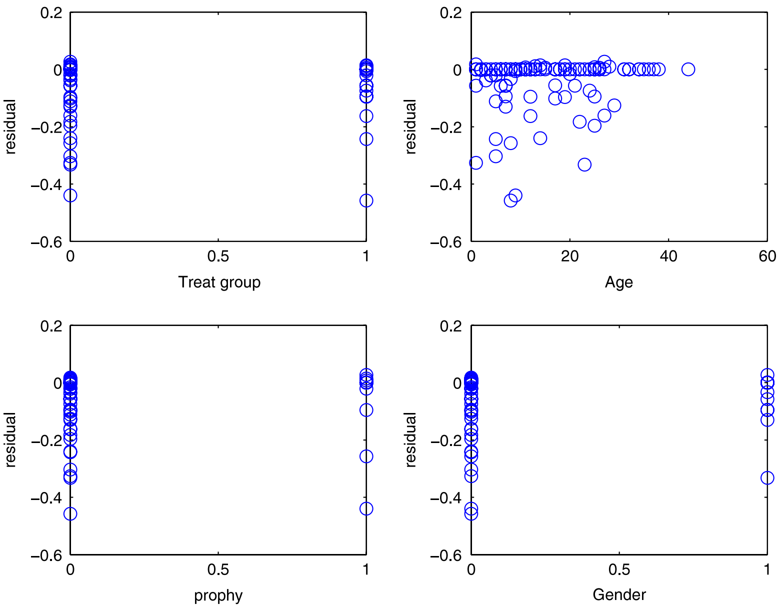

Fig. 9. Plots of residual versus covariates for the additive Cox model.

by the National Natural Science Foundation of China (Grant Nos. 11231010 and 11171330), Key Laboratory of RCSDS, CAS (No. 2008DP173182) and BCMIIS. Zhao's research was partly supported by the Research Grants Council of Hong Kong (PolyU 504011), the National Natural Science Foundation of China (No. 11371299), and The Hong Kong Polytechnic University.

Appendix A. Proof of Theorem 1

Let $s_0(t;\beta,\lambda)$, $s_z(t;\beta,\lambda)$, $e_z(t;\beta,\lambda)$, $e_{zz}(t;\beta,\lambda)$ and $\bar{z}(t;\beta,\lambda)$ be the limits of $S_0(t;\beta,\lambda)$, $S_z(t;\beta,\lambda)$, $E_z(t;\beta,\lambda)$, $E_{zz}(t;\beta,\lambda)$, and $\bar{\mathbf{Z}}(t;\beta,\lambda)$, respectively. Set $s_0(t) = s_0(t;\beta_0,\lambda_0)$, and define $s_z(t)$, $e_z(t)$, $e_{zz}(t)$, and $\bar{z}(t)$ similar to $s_0(t)$. Also let $A = \int_0^\tau e_{zz}(t)dt$.

In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

(C1) $\lambda_0(t)$ is three times continuously differentiable for $t \in [0,\tau]$.
(C2) The covariate Z is bounded.

(C3) K is a symmetric and continuous kernel function with a compact support satisfying $\int K(u)du = 1$ and $h = O(n^{-\alpha})$, where $1/8 < \alpha < 1/4$.
(C4) $s_0(t; \beta, \lambda), s_1(t; \beta, \lambda), e_z(t; \beta, \lambda)$ and $e_{zz}(t; \beta, \lambda)$ are uniformly continuous with respect to $(t; \beta, \lambda) \in [0, \tau] \times \mathcal{B} \times \mathcal{H}$, where $\mathcal{B}$ is a compact set of R^p including β_0, and $\mathcal{H}$ is a compact set of R that includes a neighborhood of $\lambda_0(t)$ for $t \in [0, \tau]$.
(C5) $s_0(t)$ and A are nonsingular for $t \in [0, \tau]$.

Proof of Theorem 1(i). To prove consistency of $\hat{\beta}$ and $\hat{\Lambda}$, we will utilize Lemmas 1 and 2 of Martinussen et al. (2002) with $R_n = cn^{-\delta}$ and $2\alpha < \delta < \frac{1}{2}$, where $c > 0$ is a constant. Define

$$Q(\beta) = \beta + \frac{1}{n}\left\{\int_0^\tau E_{zz}(t; \beta, \bar{\lambda})dt\right\}^{-1} \int_0^\tau \{\mathbf{Z} - \bar{\mathbf{Z}}(t; \beta, \bar{\lambda})\}' dM(t; \beta, \bar{\lambda}),$$

and

$$\Psi(\Lambda)(t) = \int_0^t \bar{\lambda}(u)du + \frac{1}{n}\sum_{i=1}^n \int_0^t S_0^{-1}(u; \beta, \bar{\lambda})dM_i(u; \beta, \bar{\lambda}) - \int_0^t E_z(u; \beta, \bar{\lambda})' du\{Q(\beta) - \beta\},$$

where $\bar{\lambda}(t) = \int \frac{1}{h}K\left(\frac{u-t}{h}\right) d\Lambda(u)$. Let $\bar{\lambda}_0(t) = \int \frac{1}{h}K(\frac{u-t}{h})d\Lambda_0(u)$. Then it follows from the Taylor expansion that uniformly in $t \in [0, \tau]$,

$$\bar{\lambda}_0(t) - \lambda_0(t) = O(h^2). \quad \text{(A.1)}$$

Thus, using the Taylor expansion, and the central limit theorem and the uniform strong law of large numbers (Pollard, 1990, p. 41), we have

$$\begin{aligned} Q(\beta_0) - \beta_0 &= \frac{1}{n}\left\{\int_0^\tau E_{zz}(t; \beta_0, \bar{\lambda}_0)dt\right\}^{-1} \int_0^\tau \{\mathbf{Z} - \bar{\mathbf{Z}}(t; \beta_0, \bar{\lambda}_0)\}' dM(t; \beta_0, \lambda_0) + O_p(\|\bar{\lambda}_0 - \lambda_0\|^2) \\ &= O_p\left(n^{-\frac{1}{2}}\right) + O_p(h^4). \end{aligned}$$

Therefore,

$$R_n^{-1}\|Q(\beta_0) - \beta_0\| = o_p(1),$$

and for any $\varepsilon > 0$, there exists a C such that for sufficiently large n,

$$P\Big(n^{1/2}\|Q(\beta_0) - \beta_0\| \le C\Big) > 1 - \varepsilon,$$

which implies that condition (A2) of Lemma 1 and condition (C) of Lemma 2 in Martinussen et al. (2002) are satisfied for $Q(\beta_0)$.

Note that for any $\|\beta_1 - \beta_0\| \le R_n$ and $\|\beta_2 - \beta_0\| \le R_n$,

$$Q(\beta_1) - Q(\beta_2) = \int_0^\tau \{h_t(\beta_1) - h_t(\beta_2)\}dt + \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{W_i(t; \beta_1) - W_i(t; \beta_2)\}dM_i(t; \beta_0, \lambda_0), \quad \text{(A.2)}$$

where for $k = 1$ and 2,

$$\begin{aligned} h_t(\beta_k) &= \frac{1}{n}\sum_{i=1}^n \left\{\int_0^\tau E_{zz}(t; \beta_k, \bar{\lambda})dt\right\}^{-1} \{Z_i - E_z(t; \beta_k, \bar{\lambda})\}\frac{1}{S_i}\sum_{j=1}^{S_i} Y_{ij}(t)\Big[H\{\lambda_0(t) + Z_i'\beta_0\} \\ &\quad - H\{\bar{\lambda}(t) + Z_i'\beta_k\} - \dot{H}\{\bar{\lambda}(t) + Z_i'\beta_k\}Z_i'(\beta_0 - \beta_k)\Big], \end{aligned}$$

and

$$W_i(t; \beta_k) = \left\{\int_0^\tau E_{zz}(t; \beta_k, \bar{\lambda})dt\right\}^{-1} \{Z_i - E_z(t; \beta_k, \bar{\lambda})\}.$$

It can be shown that $h_t(\beta_0) = 0$, and $h_t'(\beta_0) = \left.\frac{\partial h_t(\beta)}{\partial \beta}\right|_{\beta=\beta_0} = 0$. Then it follows from the mean value theorem that

$$\begin{aligned} \|h_t(\beta_1) - h_t(\beta_2)\| &= \|h_t'(\beta^*)(\beta_1 - \beta_2)\| \\ &= \|(h_t'(\beta^*) - h_t'(\beta_0))(\beta_1 - \beta_2)\| \\ &= \|h_t''(\beta^{**})(\beta^* - \beta_0)(\beta_1 - \beta_2)\|, \end{aligned} \quad \text{(A.3)}$$

where β^* is between β_1 and β_2, β^{**} is between β^* and β_0, and $h_t''(\cdot)$ is the second derivative of $h_t(\cdot)$. Since $\|\beta^* - \beta_0\| \leq \max\{\|\beta_1 - \beta_0\|, \|\beta_2 - \beta_0\|\} \leq R_n$ and $h_t''(\cdot)$ is bounded, we obtain that

$$\|h_t(\beta_1) - h_t(\beta_2)\| \leq O_p(R_n)\|\beta_1 - \beta_2\|. \tag{A.4}$$

Likewise, for every i, there exists a $\tilde{\beta}^*$ such that

$$W_i(t;\beta_1) - W_i(t;\beta_2) = W_i'(t;\tilde{\beta}^*)(\beta_1 - \beta_2),$$

where $W_i'(t;\beta)$ denotes the derivative of $W_i(t;\beta)$ with respect to β. Thus, by the central limit theorem, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{W_i(t;\beta_1) - W_i(t;\beta_2)\}dM_i(t;\beta_0,\lambda_0) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} W_i'(t;\tilde{\beta}^*)dM_i(t;\beta_0,\lambda_0)(\beta_1 - \beta_2)\\ &= O_p\left(n^{-\frac{1}{2}}\right)\|\beta_1 - \beta_2\|.\end{aligned} \tag{A.5}$$

Thus, it follows from (A.2), (A.4) and (A.5) that for some $\theta < 1$ and for all large n,

$$\|Q(\beta_1) - Q(\beta_2)\| \leq \theta\|\beta_1 - \beta_2\|,$$

which means that $Q(\beta)$ is a contraction map. That is, condition (A1) of Lemma 1 of Martinussen et al. (2002) is satisfied for $Q(\beta)$.

Next, we verify conditions (A1) and (A2) of Lemma 1 and condition (C) of Lemma 2 in Martinussen et al. (2002) for $\Psi(\Lambda)$. Using the Taylor expansion and (A.1), we have that uniformly in $t \in [0,\tau]$,

$$\begin{aligned}\Psi(\Lambda_0)(t) &= \int_0^t \bar{\lambda}_0(u)du + \frac{1}{n}\sum_{i=1}^{n}\int_0^t S_0^{-1}(u;\beta_0,\bar{\lambda}_0)dM_i(u;\beta_0,\bar{\lambda}_0) - \int_0^t E_z(u;\beta_0,\bar{\lambda}_0)'\{Q(\beta_0) - \beta_0\}du\\
&= \int_0^t \bar{\lambda}_0(u)du + \frac{1}{n}\sum_{i=1}^{n}\int_0^t S_0^{-1}(u;\beta_0,\bar{\lambda}_0)dM_i(u;\beta_0,\lambda_0)\\
&\quad + \frac{1}{n}\int_0^t S_0^{-1}(u;\beta_0,\bar{\lambda}_0)\sum_{i=1}^{n}\phi_i(u;\beta_0,\bar{\lambda}_0)\{\lambda_0(u) - \bar{\lambda}_0(u)\}du\\
&\quad - \int_0^t E_z(u;\beta_0,\bar{\lambda}_0)'\{Q(\beta_0) - \beta_0\}du + O_p(\|\bar{\lambda}_0 - \lambda_0\|^2)\\
&= \Lambda_0(t) + \frac{1}{n}\sum_{i=1}^{n}\int_0^t S_0^{-1}(u;\beta_0,\bar{\lambda}_0)dM_i(u;\lambda_0,\beta_0)\\
&\quad - \int_0^t E_z(u;\beta_0,\bar{\lambda}_0)'\{Q(\beta_0) - \beta_0\}du + O_p(\|\bar{\lambda}_0 - \lambda_0\|^2)\\
&= \Lambda_0(t) + O_p(n^{-1/2}) + O_p(h^4).\end{aligned}$$

Hence,

$$R_n^{-1}\|\Psi(\Lambda_0) - \Lambda_0\| = o_p(1), \tag{A.6}$$

and for any $\varepsilon > 0$, there exists a C such that for sufficiently large n,

$$P\left(n^{1/2}\|\Psi(\Lambda_0) - \Lambda_0\| \leq C\right) > 1 - \varepsilon, \tag{A.7}$$

where $\|g\| = \sup_{0\leq t\leq\tau}|g(t)|$ for any function g.

Note that for any Λ_1 and Λ_2 satisfying $\|\Lambda_1 - \Lambda_0\| \leq R_n$ and $\|\Lambda_2 - \Lambda_0\| \leq R_n$, we obtain that for given β,

$$\begin{aligned}\Psi(\Lambda_1)(t) - \Psi(\Lambda_2)(t) &= \int_0^t\{f_u(\bar{\lambda}_1(u)) - f_u(\bar{\lambda}_2(u))\}du\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^t\{S_0^{-1}(u;\beta,\bar{\lambda}_1) - S_0^{-1}(u;\beta,\bar{\lambda}_2)\}dM_i(u;\beta_0,\lambda_0),\end{aligned} \tag{A.8}$$

where for $k = 1$ and 2,

$$f_u(\bar{\lambda}_k(u)) = \frac{1}{n}\sum_{i=1}^{n} S_0^{-1}(u;\beta,\bar{\lambda}_k)\frac{1}{S_i}\sum_{j=1}^{S_i} Y_{ij}(u)\Big[H\{\lambda_0(u)+Z_i'\beta_0\} - H\{\bar{\lambda}_k(u)+Z_i'\beta\} + \dot{H}\{\bar{\lambda}_k(u)+Z_i'\beta\}\{\bar{\lambda}_k(u)-\lambda_0(u)\}\Big],$$

and

$$\bar{\lambda}_k(t) = \int \frac{1}{h}K\left(\frac{u-t}{h}\right)d\Lambda_k(u).$$

It can be shown that $f_u(\lambda_0(u)) = 0$ and $f_u'(\lambda_0(u)) = \frac{\partial f_u(\lambda)}{\partial\lambda}\Big|_{\lambda=\lambda_0(u)} = 0$. By the arguments leading to (A.3), we have

$$\|f_t(\bar{\lambda}_1) - f_t(\bar{\lambda}_2)\| \le c\|\bar{\lambda}_1-\bar{\lambda}_2\|\,\|\lambda^*-\lambda_0\|,$$

where $\|\lambda^*-\lambda_0\| \le \max\{\|\bar{\lambda}_1-\lambda_0\|, \|\bar{\lambda}_2-\lambda_0\|\}$. Note that

$$\bar{\lambda}_1(t)-\bar{\lambda}_2(t) = \int h^{-1}K\left(\frac{u-t}{h}\right)d[\Lambda_1(u)-\Lambda_2(u)] = -\int h^{-2}K_d\left(\frac{u-t}{h}\right)[\Lambda_1(u)-\Lambda_2(u)]du,$$

where K_d is the derivative of K. Thus,

$$\|\bar{\lambda}_1-\bar{\lambda}_2\| \le h^{-1}\|\Lambda_1-\Lambda_2\|.$$

In a similar manner, we can prove that $\|\bar{\lambda}_0-\bar{\lambda}_k\| \le h^{-1}\|\Lambda_0-\Lambda_k\| = O(n^{\alpha-\delta})$. It then follows from (A.1) that

$$\|\lambda_0-\bar{\lambda}_j\| = O(h^2)+O(n^{\alpha-\delta}).$$

Therefore,

$$\|f_t(\bar{\lambda}_1)-f_t(\bar{\lambda}_2)\| = [O(h)+O(n^{2\alpha-\delta})]\|\Lambda_1-\Lambda_2\|,$$

which implies the first term on the right-hand side of (A.8) is $o_p(\|\Lambda_1-\Lambda_2\|)$. By the same argument as that of (A.5), we get that the second term on the right-hand side of (A.8) is also $o_p(\|\Lambda_1-\Lambda_2\|)$. Hence for sufficiently large n,

$$\|\Psi(\Lambda_1)-\Psi(\Lambda_2)\| \le \theta\|\Lambda_1-\Lambda_2\|. \tag{A.9}$$

In view of (A.6), (A.7) and (A.9), conditions (A.1) and (A.2) of Lemma 1 and condition (C) of Lemma 2 in Martinussen et al. (2002) are satisfied for $\Psi(\Lambda)$. Thus, it follows that with probability tending to one, (6) and (7) have solutions $Q(\hat{\beta}) = \hat{\beta}$ and $\Psi(\hat{\Lambda}) = \hat{\Lambda}$ such that $\|\hat{\beta}-\beta_0\| = O_p(n^{-1/2})$ and $\|\hat{\Lambda}-\Lambda_0\| = O_p(n^{-1/2})$.

Proof of Theorem 1(ii). Note that

$$\hat{\lambda}(t)-\lambda_0(t) = \int h^{-1}K\left(\frac{u-t}{h}\right)d[\hat{\Lambda}(u)-\Lambda_0(u)] + \{\bar{\lambda}_0(t)-\lambda_0(t)\}.$$

Then it follows from condition (C3) and Theorem 1(i) that

$$\|\hat{\lambda}-\lambda_0\| \le O_p(h^{-1}\|\hat{\Lambda}-\Lambda_0\|)+O(h^2) = o_p(n^{-1/4}). \tag{A.10}$$

Using the functional central limit theorem (Pollard, 1990, p. 53), we have that uniformly in $t \in [0,\tau]$, for $k = 0$ and z,

$$\|S_k(t;\beta_0,\lambda_0)-s_k(t)\| = O_p(n^{-1/2}).$$

Applying the Taylor expansion and (A.10), we obtain

$$\sup_{0\le t\le\tau}\|S_k(t;\hat{\beta},\hat{\lambda})-S_k(t;\beta_0,\lambda_0)\| = O_p(\|\hat{\beta}-\beta_0\|)+O_p(\|\hat{\lambda}-\lambda_0\|) = o_p(n^{-1/4}).$$

Thus,

$$\sup_{0\le t\le\tau}\|S_k(t;\hat{\beta},\hat{\lambda})-s_k(t)\| = o_p(n^{-1/4}). \tag{A.11}$$

Similarly, for $k = z$ and zz,

$$\sup_{0\le t\le\tau}\|E_k(t;\hat{\beta},\hat{\lambda})-e_k(t)\| = o_p(n^{-1/4}). \tag{A.12}$$

To derive the asymptotic distribution of $\hat{\beta}$, it follows from (6) that

$$\begin{aligned} Q(\hat{\beta}) - \hat{\beta} &= \frac{1}{n}\Big[\int_0^\tau E_{zz}(t; \hat{\beta}, \hat{\lambda})dt\Big]^{-1}\int_0^\tau \{\mathbf{Z} - \bar{\mathbf{Z}}(t; \hat{\beta}, \hat{\lambda})\}'[dM(t; \hat{\beta}, \hat{\lambda}) - dM(t; \beta_0, \lambda_0)] \\ &\quad + \frac{1}{n}\Big[\int_0^\tau E_{zz}(t; \hat{\beta}, \hat{\lambda})dt\Big]^{-1}\int_0^\tau \{\mathbf{Z} - \bar{\mathbf{Z}}(t; \hat{\beta}, \hat{\lambda})\}'dM(t; \beta_0, \lambda_0) \\ &= \frac{1}{n}\sum_{i=1}^n \Big[\int_0^\tau E_{zz}(t; \hat{\beta}, \hat{\lambda})dt\Big]^{-1}\int_0^\tau \{Z_i - E_z(t; \hat{\beta}, \hat{\lambda})\} \\ &\quad \times \Big[\phi_i(t; \hat{\beta}, \hat{\lambda})\{\lambda_0(t) - \hat{\lambda}(t)\} + \phi_i(t; \hat{\beta}, \hat{\lambda})Z_i'(\beta_0 - \hat{\beta})\Big]dt \\ &\quad + \frac{1}{n}\sum_{i=1}^n \Big[\int_0^\tau E_{zz}(t; \hat{\beta}, \hat{\lambda})dt\Big]^{-1}\int_0^\tau \{Z_i - E_z(t; \hat{\beta}, \hat{\lambda})\}dM_i(t; \beta_0, \lambda_0) \\ &\quad + O_p(\|\hat{\lambda} - \lambda_0\|^2) + O_p(\|\hat{\beta} - \beta_0\|^2). \end{aligned} \tag{A.13}$$

It is easy to see that

$$\sum_{i=1}^n \{Z_i - E_z(t; \hat{\beta}, \hat{\lambda})\}\phi_i(t; \hat{\beta}, \hat{\lambda}) = 0,$$

and

$$\sum_{i=1}^n \{Z_i - E_z(t; \hat{\beta}, \hat{\lambda})\}\phi_i(t; \hat{\beta}, \hat{\lambda})Z_i(t) = nE_{zz}(t; \hat{\beta}, \hat{\lambda}).$$

Hence the first term on the right side of (A.13) is equivalent to $\beta_0 - \hat{\beta}$. Using (A.11), (A.12) and Lemma 1 of Lin et al. (2000), the second term on the right side of (A.13) can be expressed as

$$\frac{1}{n}\sum_{i=1}^n A^{-1}\int_0^\tau \{Z_i - e_z(t)\}dM_i(t; \beta_0, \lambda_0) + o_p(n^{-1/2}).$$

Note that by (A.10) and the consistency of $\hat{\beta}$, we have

$$O_p(\|\lambda_0 - \hat{\lambda}\|^2) + O_p(\|\beta_0 - \hat{\beta}\|^2) = o_p(n^{-1/2}).$$

Thus, it follows from (A.13) that

$$n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2}\sum_{i=1}^n A^{-1}\xi_i + o_p(1), \tag{A.14}$$

which is a sum of i.i.d. zero-mean random vectors plus an asymptotically negligible term, where

$$\xi_i = \int_0^\tau \{Z_i - e_z(t)\}dM_i(t; \beta_0, \lambda_0).$$

It then follows from the multivariate central limit theorem and (A.14) that $n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution to a normal random variable with mean zero and variance matrix $A^{-1}E\{\xi_i^{\otimes 2}\}A^{-1}$, which can be consistently estimated by $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$ given in Theorem 1(ii).

Proof of Theorem 1(iii). It follows from (7) that

$$\begin{aligned} \Psi(\hat{\Lambda})(t) &= \int_0^t \hat{\lambda}(u)du + \frac{1}{n}\sum_{i=1}^n \int_0^t S_0^{-1}(u; \hat{\beta}, \hat{\lambda})dM_i(u; \hat{\beta}, \hat{\lambda}) \\ &= \int_0^t \hat{\lambda}(u)du + \frac{1}{n}\sum_{i=1}^n \int_0^t S_0^{-1}(u; \hat{\beta}, \hat{\lambda})\{dM_i(u; \hat{\beta}, \hat{\lambda}) - dM_i(u; \beta_0, \lambda_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^n \int_0^t S_0^{-1}(u; \hat{\beta}, \hat{\lambda})dM_i(u; \beta_0, \lambda_0). \end{aligned} \tag{A.15}$$

By the consistency of $\hat{\beta}$ and (A.10), we have that uniformly in $t \in [0, \tau]$,

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t S_0^{-1}(u;\hat{\beta},\hat{\lambda})\{dM_i(u;\hat{\beta},\hat{\lambda}) - dM_i(u;\beta_0,\lambda_0)\}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^t S_0^{-1}(u;\hat{\beta},\hat{\lambda})\phi_i(u;\hat{\beta},\hat{\lambda})[\{\lambda_0(u)-\hat{\lambda}(u)\}du + Z_i'(\beta_0-\hat{\beta})du] + o_p\left(n^{-\frac{1}{2}}\right).$$

Thus, applying (A.12), the second term on the right side of (A.15) is equivalent to

$$\int_0^t \{\lambda_0(u)-\hat{\lambda}(u)\}du - \int_0^t e_z(u)'(\hat{\beta}-\beta_0)du + o_p(n^{-1/2}) \tag{A.16}$$

uniformly in $t \in [0, \tau]$. Using (A.11), the third term on the right side of (A.15) equals

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t s_0^{-1}(u)dM_i(u;\beta_0,\lambda_0) + o_p(n^{-1/2}) \tag{A.17}$$

uniformly in $t \in [0, \tau]$. Thus, it follows from (A.14)–(A.17) that uniformly in $t \in [0, \tau]$,

$$n^{1/2}\{\hat{\Lambda}(t) - \Lambda_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\eta_i(t) + o_p(1), \tag{A.18}$$

where

$$\eta_i(t) = \int_0^t s_0^{-1}(u)dM_i(u;\beta_0,\lambda_0) - \int_0^t e_z(u)'A^{-1}\xi_i du.$$

By the multivariate central limit theorem, $n^{1/2}\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ converges in finite dimensional distributions to a zero-mean Gaussian process. Since the first term on the right-hand side of (A.18) can be written as sums of monotone processes, it is tight by Example 2.11.16 of van der Vaart and Wellner (1996). The second term is tight because $\int_0^t e_z(u)'du$ is a deterministic function. Thus, $n^{1/2}\{\hat{\Lambda}(t)-\Lambda_0(t)\}$ is tight and converges weakly to a zero-mean Gaussian process whose covariance function at (s, t) can be consistently estimated by $\hat{\Gamma}(s, t)$ given in Theorem 1(iii).

Appendix B. Proof of (9) in Section 3

We only give the proof of (9) in Section 3. The weak convergence of $\mathcal{F}_j(\tau, z)$ can be derived similarly. Taking the linear expansion of $H(x)$ and using the same arguments used in the proof of Theorem 1(ii), we can obtain that uniformly in t and z,

$$\mathcal{F}(t,z) = n^{-1/2}\sum_{i=1}^{n}\int_0^t I\{Z_i \le z\}dM_i(u;\beta_0,\lambda_0) - \int_0^t \Phi(u,z)n^{1/2}\{\hat{\lambda}(u)-\lambda_0(u)\}du$$
$$- B(t,z)n^{1/2}(\hat{\beta}-\beta_0) + o_p(1), \tag{B.19}$$

where

$$\Phi(u,z) = E\Big[I\{Z_i \le z\}\phi_i(u;\beta_0,\lambda_0)\Big],$$

and

$$B(t,z) = E\Big[\int_0^t I\{Z_i \le z\}\phi_i(u;\beta_0,\lambda_0)Z_i du\Big].$$

Let $\tilde{\Lambda}(t) = \int_0^t \hat{\lambda}(u)du$, and $\tilde{\Lambda}_0(t) = \int_0^t \bar{\lambda}_0(u)du$. Using integration by parts, it is easy to see that $\tilde{\Lambda}(t) = \int \hat{\Lambda}(y)\frac{1}{h}K\left(\frac{y-t}{h}\right)dy$ and $\tilde{\Lambda}_0(t) = \int \Lambda_0(y)\frac{1}{h}K\left(\frac{y-t}{h}\right)dy$. In view of (A.18), by the same argument as that of Theorem 2.15 of Stute (1982), we have that for any $y, u \in [0, \tau]$,

$$\sup_{|y-u|\le ch_n} |\hat{\Lambda}(y) - \Lambda_0(y) - \hat{\Lambda}(u) + \Lambda_0(u)| = o_p(n^{-1/2}).$$

Hence, using integration by parts yields that uniformly in t and z,

$$\int_0^t \Phi(u,z)n^{1/2}d[\tilde{\Lambda}(u) - \tilde{\Lambda}_0(u) - \hat{\Lambda}(u) + \Lambda_0(u)] = o_p(1). \tag{B.20}$$

Using the Taylor expansion, we get that uniformly in t and z,

$$\int_0^t \Phi(u,z)n^{1/2}d[\tilde{\Lambda}_0(u)-\Lambda_0(u)] = O(n^{1/2}h^r) = o(1). \tag{B.21}$$

Note that the second term on the right-hand side of (B.19) can be written as

$$-\int_0^t \Phi(u,z)n^{1/2}d[\tilde{\Lambda}(u)-\tilde{\Lambda}_0(u)-\hat{\Lambda}(u)+\Lambda_0(u)] - \int_0^t \Phi(u,z)n^{1/2}d[\tilde{\Lambda}_0(u)-\Lambda_0(u)]$$
$$-\int_0^t \Phi(u,z)n^{1/2}d[\hat{\Lambda}(u)-\Lambda_0(u)].$$

Then by (B.20) and (B.21), the second term on the right-hand side of (B.19) equals

$$-\int_0^t \Phi(u,z)n^{1/2}d[\hat{\Lambda}(u)-\Lambda_0(u)] + o_p(1) \tag{B.22}$$

uniformly in t and z. Thus, it follows from (A.14), (A.18), (B.19) and (B.22) that uniformly in t and z,

$$\mathcal{F}(t,z) = n^{-1/2}\sum_{i=1}^{n}\Big[\int_0^t I\{Z_i \le z\}dM_i(u;\beta_0,\lambda_0) - \int_0^t \Phi(u,z)d\hat{\eta}_i(u) - B(t,z)n^{1/2}A^{-1}\xi_i + o_p(1)\Big].$$

The multivariate central limit theorem implies that $\mathcal{F}(t,z)$ converges in finite-dimensional distribution to a zero-mean Gaussian process. By the same argument as the tightness of $n^{1/2}\{\hat{\Lambda}(t)-\Lambda_0(t)\}$, $\mathcal{F}(t,z)$ is tight. Therefore, $\mathcal{F}(t,z)$ converges weakly to a zero-mean Gaussian process which can be approximated by the zero-mean Gaussian process $\tilde{\mathcal{F}}(t,z)$ given in (9).

References

Andersen, P.K., Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. Ann. Statist. 10, 1100–1120.
Chang, S.H., 2004. Estimating marginal effects in accelerated failure time models for serial sojourn among repeated event. Lifetime Data Anal. 10, 175–190.
Chen, K., Jin, Z., Ying, Z., 2002. Semiparametric analysis of transformation models with censored data. Biometrika 89, 659–668.
Cook, R.J., Lawless, J.F., 2007. The Statistical Analysis of Recurrent Events. Springer, New York.
Darlington, G.A., Dixon, S.N., 2013. Event-weighted proportional hazards modelling for recurrent gap time data. Stat. Med. 32, 124–130.
Du, P., 2009. Nonparametric modeling of the gap time in recurrent event data. Lifetime Data Anal. 15, 256–277.
Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.
Gail, M.H., Santner, T.J., Brown, C.C., 1980. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. Biometrics 36, 255–266.
Huang, Y., Chen, Y.Q., 2003. Marginal regression of gaps between recurrent events. Lifetime Data Anal. 9, 293–303.
Huang, X., Liu, L., 2007. A joint frailty model for survival and gap times between recurrent events. Biometrics 63, 389–397.
Lawless, J.F., Nadeau, C., 1995. Some simple robust methods for the analysis of recurrent events. Technometrics 37, 158–168.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Lin, D.Y., Sun, W., Ying, Z., 1999. Nonparametric estimation of the gap time distribution for serial events with censored data. Biometrika 86, 59–70.
Lin, D.Y., Wei, L.J., Yang, L., Ying, Z., 2000. Semiparametric regression for the mean and rate function of recurrent events. J. R. Stat. Soc. Ser. B 69, 711–730.
Lin, D.Y., Wei, L.J., Ying, Z., 1993. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 80, 557–572.
Lu, W., 2005. Marginal regression of multivariate event times based on linear transformation models. Lifetime Data Anal. 11, 389–404.
Luo, X., Huang, C.Y., 2011. Analysis of recurrent gap time data using the weighted risk-set method and the modified within-cluster resampling method. Stat. Med. 30, 301–311.
Martinussen, T., Scheike, T.H., Skovgaard, I.M., 2002. Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. Scand. J. Statist. 29, 57–74.
Murphy, S.A., 1995. Asymptotic theory for the frailty model. Ann. Statist. 23, 182–198.
Nielsen, G.G., Gill, R.D., Andersen, P.K., Sorensen, T.I.A., 1992. A counting process approach to maximum likelihood estimation in frailty models. Scand. J. Statist. 19, 25–43.
Peña, E.A., Strawderman, R., Hollander, M., 2001. Nonparametric estimation with recurrent event data. J. Amer. Statist. Assoc. 96, 1299–1315.
Pollard, D., 1990. Empirical Processes: Theory and Applications. Institute of Mathematical Statistics, Hayward, CA.
Prentice, R.L., Williams, B.J., Peterson, A.V., 1981. On the regression analysis of multivariate failure time data. Biometrika 68, 373–379.
Schaubel, D.E., Cai, J., 2004. Regression analysis for gap time hazard functions of sequentially ordered multiplivariate failure time data. Biometrika 91, 291–303.
Schaubel, D.E., Zeng, D., Cai, J., 2006. A semiparametric additive rates model for recurrent event data. Lifetime Data Anal. 12, 389–406.
Strawderman, R.L., 2005. The accelerated gap times model. Biometrika 92, 1299–1315.
Stute, W., 1982. The oscillation behavior of empirical processes. Ann. Statist. 10, 86–107.
Sun, L., Park, D.H., Sun, J., 2006. The additive hazards model for recurrent gap times. Statist. Sinica 16, 919–932.
Sun, L., Tong, X., Zhou, X., 2011. A class of Box–Cox transformation models for recurrent event data. Lifetime Data Anal. 17, 280–301.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer, New York.
Wand, M.P., Schucany, W.R., 1990. Gaussian-based kernels. Canad. J. Statist. 18, 197–204.
Wang, M.C., Chang, S.H., 1999. Nonparametric estimation of a recurrent survival function. J. Amer. Statist. Assoc. 94, 146–153.
Zeng, D., Lin, D.Y., 2006. Efficient estimation of semiparametric transformation models with random effects for recurrent events. Biometrika 93, 627–640.
Zeng, D., Lin, D.Y., 2007. Semiparametric transformation models for counting processes. J. Amer. Statist. Assoc. 102, 167–180.
Zeng, D., Yin, G., Ibrahim, J.G., 2005. Inference for a class of transformed hazards model. J. Amer. Statist. Assoc. 100, 1000–1008.