The Additive Hazards Model for Multiple Type Recurrent Gap Times *

Liu Jicai[1]　Zhang Riquan[1,2]　Liu Huanbin[3]

([1] *School of Finance and Statistics, East China Normal University, Shanghai, 200241*)
([2] *Department of Mathematics, Shanxi Datong University, Datong, 037009*)
([3] *College of Mathematics and Computer Science, Huanggang Normal University, Huanggang, 438000*)

Abstract

In many biomedical and engineering studies, recurrent event data and gap times between successive events are common and often more than one type of recurrent events is of interest. It is well known that the proportional hazards model may not be appropriate for fitting survival times in some settings. In the paper, we consider an additive hazards model for multiple type recurrent gap times data to assess the effect of covariates. For inferences about regression coefficients and baseline cumulative hazard functions, an estimating equation approach is developed. Furthermore, we establish asymptotic properties of the proposed estimators.

Keywords: Additive hazards model, gap times, multiple type recurrent events, estimating equation, multivariate survival analysis.

AMS Subject Classification: 62N01, 62G05.

§1. Introduction

Recurrent event data arise frequently in many research areas when events of interest can occur repeatedly over time for each subject. Examples of such recurrent event data include infection occurrences among patients receiving transplants, bladder tumor recurrences, repeated purchases of a particular type of certain product and repeated failures of a certain machine. Moreover, in many settings, several different but related types of recurrent events may occur together and in these cases one faces multiple type recurrent event data. Many authors have investigated the analysis of recurrent event data. For example, for univariate recurrent event data, Prentice et al. (1981) and Anderson and Gill

*The second author was supported in part by National Natural Science Foundation of China (1117112), Doctoral Fund of Ministry of Education of China (20090076110001) and National Statistical Science Research Major Program of China (2011LZ051). The third author was supported in part by Natural Science Foundation of Hubei (2011CDB167) and Major Research Program of Hubei Provincial Department of Education (Z20092701).

Received October 8, 2012. Revised January 22, 2013.

(1982) proposed conditional model methods while Wei et al. (1989) and Pepe and Cai (1993) developed marginal model methods. The authors who considered the analysis of multiple type recurrent event data include Spiekerman and Lin (1998), Clegg et al. (1999) and Cai and Schaubel (2004).

In many applications, investigators are also interested in the gap time between successive events. A number of methods have been proposed for analyzing such data. For example, Lin et al. (1999), Wang and Chang (1999) and Peña et al. (2001) have developed nonparametric methods to estimate the distribution of the gap times; Huang and Chen (2003) and Sun et al. (2006) proposed the proportional hazards model and the additive hazards model based on a renewal process to evaluate covariate effects, respectively. However, to our knowledge, there are few results for multiple type recurrent gap times data. This paper focuses on the statistical analysis of such data.

To describe multiple type recurrent gap times data, suppose that there are K different types of recurrent events of interest and a total of n subjects are observed. Let T_{ikj} denote the time from the $(j-1)$th occurrence to the jth occurrence of the event of type k, where $i=1,\ldots,n$; $j=1,2,\ldots$ and $k=1,\ldots,K$. That is, $T_{ik1}+T_{ik2}+T_{ik1}+\cdots+T_{ikj}$ is the time at which the event of type k occurs for the jth time of the ith subject. Denote the time-independent covariates and the censoring time by Z_{ik} and C_{ik}, respectively.

Suppose that $\{(T_{ik},Z_{ik},C_{ik}),i=1,\ldots,n\}$ are n i.i.d. replicates of (T_k,Z_k,C_k) for the event of type k, where $T_{ik}=\{T_{ikj}:j=1,2,\ldots\}$. We also assume that T_{ik} is independent of C_{ik}, given Z_{ik}. Define an integer M_{ik}, satisfying

$$\sum_{j=1}^{M_{ik}-1} T_{ikj}\le C_{ik} \qquad \text{and} \qquad \sum_{j=1}^{M_{ik}} T_{ikj}> C_{ik}.$$

The observed data are $(T_{ik1},\ldots,T_{ikM_{ik}-1};Z_{ik};C_{ik})$. That is, the first $M_{ik}-1$ gap times are completely observed, but $T_{ikM_{ik}}$ is censored at $T^{+}_{ikM_{ik}}=C_{ik}-\sum\limits_{j=1}^{M_{ik}-1}T_{ikj}$. Denote $\Delta_{ik}=I(M_{ik}>1)$ and $M^*_{ik}=\max(M_{ik}-1,1)$. Let

$$X_{ikj}=\begin{cases} T_{ikj} & \text{if } \ \Delta_{ik}=1;\\ T^{+}_{ikj} & \text{if } \ \Delta_{ik}=0,\end{cases} \tag{1.1}$$

for $j=1,\ldots,M^*_{ik}$, where $I(\cdot)$ is the indicator function.

In some settings, the proportional hazards model may not fit the data very well. Lin and Ying (1994) investigated the additive hazards model

$$\lambda(t|Z_i)=\lambda_0(t)+\beta_0^T Z_i, \tag{1.2}$$

where $\lambda_0(t)$ is a baseline hazard function and β_0 is an unspecified vector of parameter of interest.

For single type recurrent gap times, Sun et al. (2006) proposed an estimating equation $U^*(\beta)=0$ of parameters β_0 in model (1.2), where

$$\begin{aligned} U^*(\beta) &= \int_0^\tau Q(t)\Big\{\widehat{\mathcal{E}}_{ij}\{Z_i\Delta_i \mathrm{d}I(X_{ij}\le t)\} - \frac{\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\ge t)Z_i\}}{\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\ge t)\}}\mathrm{d}\widehat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij}\le t)\} \\ &\quad -\Big[\widehat{\mathcal{E}}_{ij}\{Z_i^{\otimes 2}I(X_{ij}\ge t)\} - \frac{(\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\ge t)Z_i\})^{\otimes 2}}{\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\ge t)\}}\Big]\beta \mathrm{d}t\Big\}, \end{aligned}$$

with $\widehat{\mathcal{E}}_{ij}=\widehat{\mathcal{E}}_i\widehat{\mathcal{E}}_j$, where $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{E}}_j$ denote empirical averages over $i=1,\ldots,n$ and $j=1,\ldots,M_i^*$, respectively; X_{ij} and M_i^* is defined similarly as (1.1). $Q(t)$ is a weight process.

Let $\widehat{\beta}^*$ denote the solution to $U^*(\beta)=0$. The baseline cumulative hazard function $\Lambda_0(t)=\int_0^t\lambda_0(s)\mathrm{d}s$ can be estimated by $\widehat{\Lambda}_0(t,\widehat{\beta}^*)$, where

$$\widehat{\Lambda}_0(t,\widehat{\beta}^*)=\int_0^t\frac{\mathrm{d}\widehat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij}\le s)\}-\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\le s)\widehat{\beta}^{*T}Z_i\mathrm{d}s\}}{\widehat{\mathcal{E}}_{ij}\{I(X_{ij}\ge s)\}}.$$

In this article, we extend the aforementioned method to the additive hazards model for multiple type recurrent gap times. A similar estimating equation is proposed for estimation of the regression parameters and the cumulative hazards functions. The resultant estimators are proven to be consistent and asymptotically normal.

In the next section, we present the model and the corresponding inference procedures for multiple type recurrent gap times. The asymptotic properties of the proposed estimators are established in Section 3. In Section 4, some remarks are given. The technical proofs are contained in Appendix.

§2. Model and Methods

For multiple type recurrent gap times data, we consider the following additive hazards model

$$\lambda_k(t|Z_{ik})=\lambda_{0k}(t)+\beta_0^TZ_{ik},\qquad k=1,2,\ldots,K, \tag{2.1}$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function and β_0 is a $p\times 1$ vector of unknown parameters of interest.

To present the inference procedure for the unknown parameters β_0 and $\Lambda_{0k}(t)=\int_0^t\lambda_{0k}(s)\mathrm{d}s$, we need the following assumptions, which are similar to those in Huang and Chen (2003). Suppose that $\{(X_{ikj}, j = 1, \ldots, M^*_{ik}; \Delta_{ik}; Z_{ik}), i = 1, \ldots, n\}$ are n i.i.d. replicates of $\{X_{k(j)}, j = 1, \ldots, M^*_k; \Delta_k; Z_k\}$, for $k = 1, \ldots, K$. We also assume that each individual recurrent process of the specific event recurrent event process is a renewal process. That is, for given i and k, T_{ikj}, $j = 1, 2, \ldots$ are i.i.d. This implies that given C_{ik}, M_{ik} and $T^+_{ikM_{ik}}$, the observed complete gap times $\{T_{ikj}, j = 1, \ldots, M_{ik} - 1\}$ are identically distributed. Since the first gap time is subject to independent censorship, the exchangeability of observed complete gap times then suggests that we can treat the subset $\{(X_{ikj}, j = 1, \ldots, M^*_{ik}; \Delta_{ik}; Z_{ik}), i = 1, \ldots, n\}$ as clustered survival data. But, the cluster size is informative and the censored gap time $T_{ikM_{ik}}$ is removed for $M_{ik} > 1$.

Following the methods proposed in Cai and Schaubel (2004) and Sun et al. (2006), we can formulate the estimating equation $U(\beta) = 0$ of the parameters β_0, where

$$\begin{aligned} U(\beta) = & \sum_{k=1}^{K} \int_0^{\tau} Q(t) \Big[\widehat{\mathcal{E}}_{ij_k}\{Z_{ik}\Delta_{ik} \mathrm{d}I(X_{ikj} \le t)\} \\ & - \frac{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t) Z_{ik}\}}{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t)\}} \mathrm{d}\widehat{\mathcal{E}}_{ij_k}\{\Delta_{ik} I(X_{ikj} \le t)\} \\ & - \widehat{\mathcal{E}}_{ij_k}\{Z^{\otimes 2}_{ik} I(X_{ikj} \ge t)\}\beta \mathrm{d}t + \frac{(\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t) Z_{ik}\})^{\otimes 2}}{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t)\}} \beta \mathrm{d}t \Big], \end{aligned} \tag{2.2}$$

with $\widehat{\mathcal{E}}_{ij_k} = \widehat{\mathcal{E}}_i \widehat{\mathcal{E}}_{j_k}$, where $\widehat{\mathcal{E}}_i$ denotes empirical averages over $i = 1, \ldots, n$ and $\widehat{\mathcal{E}}_{j_k}$ represents the empirical averages over $j = 1, \ldots, M^*_{ik}$ for the type k of event. The constant τ in $(0, \infty)$ is the study ending time, which satisfies $\mathsf{P}(X_{k(1)} \ge \tau) > 0$. $Q(t)$ is a weight process that may depend on data.

If we only use the time to the first occurrence of each event $\{(X_{ik1}; \Delta_{ik}; Z_{ik}), i = 1, \ldots, n\}$, an alternative but less efficient estimator of β_0 is obtained. The estimating equation $U_1(\beta)$ is defined by

$$\begin{aligned} U_1(\beta) = & \sum_{k=1}^{K} \int_0^{\tau} Q(t) \Big[\widehat{\mathcal{E}}_i\{Z_{ik}\Delta_{ik} \mathrm{d}I(X_{ik1} \le t)\} \\ & - \frac{\widehat{\mathcal{E}}_i\{I(X_{ik1} \ge t) Z_{ik}\}}{\widehat{\mathcal{E}}_i\{I(X_{ik1} \ge t)\}} \mathrm{d}\widehat{\mathcal{E}}_i\{\Delta_{ik} I(X_{ik1} \le t)\} \\ & - \widehat{\mathcal{E}}_i\{Z^{\otimes 2}_{ik} I(X_{ik1} \ge t)\}\beta \mathrm{d}t + \frac{(\widehat{\mathcal{E}}_i\{I(X_{ik1} \ge t) Z_{ik}\})^{\otimes 2}}{\widehat{\mathcal{E}}_i\{I(X_{ik1} \ge t)\}} \beta \mathrm{d}t \Big]. \end{aligned}$$

Denote $\widehat{L}_k(t) = \widehat{\mathcal{E}}_{ij_k}\{\Delta_{ik} I(X_{ikj} \le t)\}$, $\widehat{G}_{0k}(t) = \widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t)\}$ and $\widehat{G}_{1k}(t) = \widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj} \ge t) Z_{ik}\}$. Let $\widehat{\beta}$ be the solution to $U(\beta) = 0$. By some simple algebra manipulation, $\widehat{\beta}$ has the following closed form

$$\begin{aligned}\widehat{\beta} \;=\; & \Big\{\sum_{k=1}^K \int_0^\tau Q(t)\Big(\widehat{\mathcal{E}}_{ij_k}\{Z_{ik}^{\otimes 2} I(X_{ikj}\geq t)\} - \frac{(\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq t)Z_{ik}\})^{\otimes 2}}{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq t)\}}\Big)\mathrm{d}t\Big\}^{-1} \\ & \times\Big[\sum_{k=1}^K \int_0^\tau Q(t)\Big(\widehat{\mathcal{E}}_{ij_k}\{Z_{ik}\Delta_{ik}\mathrm{d}I(X_{ikj}\leq t)\} \\ & - \frac{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq t)Z_{ik}\}}{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq t)\}}\,\mathrm{d}\widehat{\mathcal{E}}_{ij}\{\Delta_{ik} I(X_{ikj}\leq t)\}\Big)\Big].\end{aligned}$$

Let $\widehat{\beta}_1$ be the solution to $U_1(\beta)=0$. Similar to $\widehat{\beta}$, $\widehat{\beta}_1$ also has a closed form. As expected, Theorem 3.3 in Section 3 shows that $\widehat{\beta}$ is more efficient than $\widehat{\beta}_1$. That is, the asymptotic variance of $\widehat{\beta}$ is smaller than that of $\widehat{\beta}_1$.

For the event of type k, we can obtain the Breslow-Aalen type estimator of $\Lambda_{0k}(t)$, given by

$$\widehat{\Lambda}_{0k}(t,\widehat{\beta}) = \int_0^t \frac{\mathrm{d}\widehat{\mathcal{E}}_{ij_k}\{\Delta_{ik} I(X_{ikj}\leq u)\} - \widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq u)\widehat{\beta}^T Z_{ik}\}}{\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\geq u)\}}.$$

If the first occurrence of type-k event is only used, an alternative but less efficient estimator of $\Lambda_{0k}(t)$ is

$$\widehat{\Lambda}_{0k}^{(1)}(t,\widehat{\beta}_1) = \int_0^t \frac{\mathrm{d}\hat{\mathcal{E}}_i\{\Delta_{ik} I(X_{ik1}\leq u)\} - \widehat{\mathcal{E}}_i\{I(X_{ik1}\geq u)\widehat{\beta}_1^T Z_{ik}\}}{\widehat{\mathcal{E}}_i\{I(X_{ik1}\geq u)\}}.$$

Similarly, we can show that $\widehat{\Lambda}_{0k}(t,\widehat{\beta})$ is more efficient than $\widehat{\Lambda}_{0k}^{(1)}(t,\widehat{\beta}_1)$.

§3. Asymptotic Properties

In this section, we establish the asymptotic properties of the proposed estimators. For simplicity, denote $L_k(t) = \mathsf{E}\{\Delta_k I(X_{k(1)}\leq t)\}$, $G_{0k}(t,\beta) = \mathsf{E}\{I(X_{k(1)}\geq t)\}$ and $G_{1k}(t,\beta) = \mathsf{E}\{I(X_{k(1)}\geq t)Z_k\}$. We need first to assume that the following regularity conditions hold:

(C1) $Q(t)$ has bounded variation and converges almost surely to a nonrandom function $q(t)$ uniformly over $t\in[0,\tau]$.

(C2) For each k, $\mathsf{E}\|Z_k\|\leq M$, where M is a nonrandom constant.

(C3) A is nonsingular, where

$$A = \sum_{k=1}^K \int_0^\tau q(t)\Big(\mathsf{E}\{Z_k^{\otimes 2} I(X_{k(1)}\geq t)\} - \frac{G_{1k}(t)^{\otimes 2}}{G_{0k}(t)}\Big)\mathrm{d}t.$$

To establish the asymptotic properties of $\widehat{\beta}$, we need first to investigate the asymptotic properties of $n^{1/2}U(\beta_0)$.

Theorem 3.1 Under the conditions (C1)-(C2), $n^{1/2}U(\beta_0)$ is asymptotically normal with zero mean and covariance matrix

$$\Sigma = \mathsf{E}\Big[\Big(\sum_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\}\Big)^{\otimes 2}\Big],$$

where $v^{\otimes 2} = vv^T$ for a column vector v, and

$$\begin{aligned}\Phi(X_{ikj},\Delta_{ik},Z_{ik}) &= \int_0^\tau q(t)\Big(Z_{ik} - \frac{G_{1k}(t)}{G_{0k}(t)}\Big)\Big[\Delta_{ik}\mathrm{d}I(X_{ikj}\le t) - \frac{I(X_{ikj}\ge t)}{G_{0k}(t)}\,\mathrm{d}L_k(t)\\ &\quad - I(X_{ikj}\ge t)\beta_0^T\Big(Z_{ik} - \frac{G_{1k}(t)}{G_{0k}(t)}\Big)\mathrm{d}t\Big].\end{aligned}$$

The proof is provided in Appendix. By the uniform strong law of large numbers (Pollard, 1990), the covariance matrix Σ can be consistently estimated by

$$\widehat{\Sigma} = \widehat{\mathcal{E}}_i\Big\{\Big(\sum_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\widehat{\Phi}(X_{ikj},\Delta_{ik},Z_{ik})\}\Big)^{\otimes 2}\Big\}$$

and

$$\begin{aligned}\widehat{\Phi}(X_{ikj},\Delta_{ik},Z_{ik}) &= \int_0^\tau Q(t)\Big(Z_{ik} - \frac{\widehat{G}_{1k}(t)}{\widehat{G}_{0k}(t)}\Big)\Big\{\Delta_{ik}\mathrm{d}I(X_{ikj}\le t) - \frac{I(X_{ikj}\ge t)}{\widehat{G}_{0k}(t)}\,\mathrm{d}\widehat{L}_k(t)\\ &\quad - I(X_{ikj}\ge t)\widehat{\beta}^T\Big(Z_{ik} - \frac{\widehat{G}_{1k}(t)}{\widehat{G}_{0k}(t)}\Big)\mathrm{d}t\Big\}.\end{aligned}$$

Theorem 3.2 Under the conditions (C1)-(C3), $\widehat{\beta}$ is a consistent estimator of β_0. Moreover, $n^{1/2}(\widehat{\beta}-\beta_0) \xrightarrow{D} \mathcal{N}(0, A^{-1}\Sigma A^{-1})$, where $\xrightarrow{D}$ denotes convergence in distribution.

The asymptotic variance matrix $A^{-1}\Sigma A^{-1}$ can be consistently estimated by $\widehat{\Omega} = \widehat{A}^{-1}\widehat{\Sigma}\widehat{A}^{-1}$, where

$$\widehat{A} = \sum_{k=1}^{K}\int_0^\tau Q(t)\Big(\widehat{\mathcal{E}}_{ij_k}\{Z_{ik}^{\otimes 2}I(X_{ikj}\ge t)\} - \frac{(\widehat{\mathcal{E}}_{ij_k}\{I(X_{ikj}\ge t)Z_{ik}\})^{\otimes 2}}{\widehat{\mathcal{E}}_{ij}\{I(X_{ikj}\ge t)\}}\Big)\mathrm{d}t.$$

The following Theorem 3.3 shows the asymptotic variance of $\widehat{\beta}$ is smaller than that of $\widehat{\beta}_1$.

Theorem 3.3 If the conditions (C1)-(C3) hold. Then $\widehat{\beta}$ is more efficient than $\widehat{\beta}_1$.

The asymptotic properties of the estimators of the cumulative baseline hazards functions $\Lambda_0(t) = \int_0^t \lambda_0(s)\mathrm{d}s$ for $k \in \{1, 2, \ldots, K\}$, are summarized in the following theorem.

Theorem 3.4 Under the conditions (C1)-(C3), $\widehat{\Lambda}_{0k}(t,\widehat{\beta})$ is consistent to $\Lambda_{0k}(t)$ uniformly in $t \in [0,\tau]$, i.e., $\sup_{t\in[0,\tau]} |\widehat{\Lambda}_{0k}(t,\widehat{\beta}) - \Lambda_{0k}(t)| \to 0$ almost surely. Moreover, $\sqrt{n}(\widehat{\Lambda}_{0k}(t,\widehat{\beta}) - \Lambda_{0k}(t))$ converges weakly to a zero-mean Gaussian process with covariance function

$$\Gamma_k(s,t) = \mathsf{E}\big[\widehat{\mathcal{E}}_{j_k}\{\Psi_k(s,X_{kj},\Delta_k,Z_k)\}\widehat{\mathcal{E}}_{j_k}\{\Psi_k(t,X_{kj},\Delta_k,Z_k)\}\big],$$

where

$$\begin{aligned}\Psi_k(t,X_{ikj},\Delta_{ik},Z_{ik}) \quad = \quad & -C_k^T(t)A^{-1}\Big(\sum_{m=1}^{K}\widehat{\mathcal{E}}_{jm}\{\Phi(t,X_{imj},\Delta_{im},Z_{im})\}\Big)\\ & + \int_0^t \frac{\mathrm{d}\Delta_{ik}I(X_{ikj}\le u) - I(X_{ikj}\ge u)\beta_0^T Z_{ik}\mathrm{d}u}{G_{0k}(u)}\\ & - \int_0^t \frac{I(X_{ikj}\ge u)\mathrm{d}L_k(u) - I(X_{ikj}\ge u)\beta_0^T G_{1k}(u)\mathrm{d}u}{G_{0k}(u)^2}\end{aligned}$$

and

$$C_k(t) = \int_0^t \frac{G_{1k}(u)}{G_{0k}(u)}\mathrm{d}u.$$

The covariance function can be consistently estimated by replacing limiting quantities with their empirical counterparts, i.e.,

$$\widehat{\Gamma}_k(s,t) = \widehat{\mathcal{E}}_i\{\widehat{\mathcal{E}}_{j_k}\{\widehat{\Psi}_k(s,X_{ikj},\Delta_{ik},Z_{ik})\}\widehat{\mathcal{E}}_{j_k}\{\widehat{\Psi}_k(t,X_{ikj},\Delta_{ik},Z_{ik})\}\},$$

for $k \in \{1,2,\ldots,K\}$, where

$$\begin{aligned}\widehat{\Psi}_k(t,X_{ikj},\Delta_{ik},Z_{ik}) \quad = \quad & -\widehat{C}_k^T(t)\widehat{A}^{-1}\sum_{m=1}^{K}\widehat{\mathcal{E}}_{jm}\{\widehat{\Phi}(t,X_{imj},\Delta_{im},Z_{im})\}\\ & + \int_0^t \frac{\mathrm{d}\Delta_{ik}I(X_{ikj}\le u) - I(X_{ikj}\ge u)\widehat{\beta}^T Z_{ik}\mathrm{d}u}{\widehat{G}_{0k}(u)}\\ & - \int_0^t \frac{I(X_{ikj}\ge u)\mathrm{d}\widehat{L}_k(u) - I(X_{ikj}\ge u)\widehat{\beta}^T \widehat{G}_{1k}(u)\mathrm{d}u}{\widehat{G}_{0k}(u)^2}\end{aligned}$$

and

$$\widehat{C}_k(t) = \int_0^t \frac{\widehat{G}_{1k}(u)}{\widehat{G}_{0k}(u)}\mathrm{d}u.$$

The asymptotic properties of $\widehat{\beta}_1$ and $\widehat{\Lambda}_{0k}^{(1)}(t,\widehat{\beta}_1)$ can similarly obtained, so we omit them.

§4. Concluding Remarks

In this paper, we have studied an additive hazards model for multiple type recurrent gap times. An estimating equation approach is used to estimate the regression parameters and the cumulative hazard functions. We establish asymptotic properties of the proposed estimators.

The additive mean model is another common and useful model for recurrent event data. The proposed method in this paper can be extended to the additive mean model. Our method is motivated by Cai and Schaubel (2004) and Sun et al. (2006). Thus, this extension is simple by using an estimating equation similar to one in Cai and Schaubel (2004) or (2.2).

In this paper, we consider time-independent covariates and the proposed inference approach is based on the assumption that gap times are exchangeable. That is, the observed complete gap times are identically distributed. When the covariates are time-dependent, the proposed inference approach is still applicable, as long as the gap times are assumed to be exchangeable. However, difficulties arise if the exchangeability is no longer available. More discussions on the exchangeability can be found in Wang and Chang (1999). In addition, how to choose of a weight process $Q(t)$ to obtain the most efficient estimator for β is a challenging problem. Further researches are needed to select the weight process $Q(t)$.

Appendix: Proofs of Asymptotic Properties

Proof of Theorem 3.1 By Taylor expansion, we have

$$\begin{aligned}
n^{1/2}U(\beta_0) &= n^{1/2}\sum_{k=1}^{K}\Big\{\int_0^{\tau}q(t)\Big[\widehat{\mathcal{E}}_{ij_k}\{\Delta_{ik}Z_{ik}\mathrm{d}I(X_{ikj}\le t)\}-\frac{G_{1k}(t)}{G_{0k}(t)}\mathrm{d}\widehat{L}_k(t)\\
&\quad-\frac{\widehat{G}_{1k}(t)}{G_{0k}(t)}\mathrm{d}L_k(t)+\frac{G_{1k}(t)\widehat{G}_{0k}(t)}{G_{0k}(t)^2}\mathrm{d}L_k(t)-\widehat{\mathcal{E}}_{ij_k}\{Z_{ik}^{\otimes 2}I(X_{ikj}\ge t)\}\beta_0\mathrm{d}t\\
&\quad+\Big(\frac{\widehat{G}_{1k}(t)G_{1k}(t)^T}{G_{0k}(t)}+\frac{G_{1k}(t)\widehat{G}_{1k}(t)^T}{G_{0k}(t)}-\frac{G_{1k}(t)^{\otimes 2}\widehat{G}_{0k}(t)}{G_{0k}(t)^2}\Big)\beta_0\mathrm{d}t\Big]\Big\}+o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{ikj},\Delta_{ik},Z_{ik})\}+o_p(1).
\end{aligned}\tag{A.1}$$

Note that $\sum\limits_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{ikj},\Delta_{ik},Z_{ik})\}$ are i.i.d. zero-mean random vectors for $i=1,\ldots,n$. By the multivariate central limit theorem, $n^{1/2}U(\beta_0)$ converges in distribution to a normal variable with mean zero and variance matrix Σ, which can be consistently estimated by $\widehat{\Sigma}=\widehat{\mathcal{E}}_i\Big\{\Big(\sum\limits_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\widehat{\Phi}(X_{ikj},\Delta_{ik},Z_{ik})\}\Big)^{\otimes 2}\Big\}$. □

Proof of Theorem 3.2 By the uniform strong law of large numbers, it can be show that $\widehat{A}$ converges to A and $U(\beta_0)\to 0$ almost surely. Since $\widehat{\beta}-\beta_0=\widehat{A}^{-1}U(\beta_0)$, $\widehat{\beta}$ is consistent to β_0. Note that

$$n^{1/2}(\widehat{\beta}-\beta_0)=A^{-1}n^{1/2}U(\beta_0)+o_p(1). \tag{A.2}$$

Combining Theorem 3.1, we have $n^{1/2}(\widehat{\beta}-\beta_0)\xrightarrow{D}\mathcal{N}(0,A^{-1}\Sigma A^{-1})$. The covariance matrix $A^{-1}\Sigma A^{-1}$ can be consistently estimated by $\widehat{\Omega}$. □

Proof of Theorem 3.3 Under the conditions (C1)-(C3), as for $\widehat{\beta}$, we can show that the asymptotic variance of $\widehat{\beta}_1$ is

$$A^{-1}\mathsf{E}\Big[\Big(\sum_{k=1}^{K}\Phi(X_{k(1)},\Delta_k,Z_k)\Big)^{\otimes 2}\Big]A^{-1}.$$

For each k, since

$$\begin{aligned}&\widehat{\mathcal{E}}_{j_k}\{(\Phi(X_{kj},\Delta_k,Z_k)-\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\})^{\otimes 2}\}\\ =\ &\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)^{\otimes 2}\}-(\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\})^{\otimes 2},\end{aligned}$$

we obtain

$$\mathsf{E}[(\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\})^{\otimes 2}]\le\mathsf{E}[\Phi(X_{k(1)},\Delta_{ik},Z_{ik})^{\otimes 2}]. \tag{A.3}$$

Note that $\Sigma=\mathsf{E}\Big[\Big(\sum\limits_{k=1}^{K}\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\}\Big)^{\otimes 2}\Big]$. By (A.3), we have

$$\begin{aligned}\Sigma\ =\ &\mathsf{E}\Big[\sum_{k_1\neq k_2}\widehat{\mathcal{E}}_{j_{k_1}}\{\Phi(X_{k_1j},\Delta_{k_1},Z_{k_1})\}\widehat{\mathcal{E}}_{j_{k_2}}\{\Phi(X_{k_2j},\Delta_{k_2},Z_{k_2})^T\}\Big]\\ &+\mathsf{E}\Big[\sum_{k=1}^{K}(\widehat{\mathcal{E}}_{j_k}\{\Phi(X_{kj},\Delta_k,Z_k)\})^{\otimes 2}\Big]\\ \le\ &\mathsf{E}\Big[\sum_{k_1\neq k_2}\Phi(X_{k_1(1)},\Delta_{k_1},Z_{k_1})\Phi(X_{k_2(1)},\Delta_{k_2},Z_{k_2})^T\Big]+\mathsf{E}\Big[\sum_{k=1}^{K}(\Phi(X_{k(1)},\Delta_k,Z_k))^{\otimes 2}\Big]\\ =\ &\mathsf{E}\Big[\Big(\sum_{k=1}^{K}\Phi(X_{k(1)},\Delta_k,Z_k)\Big)^{\otimes 2}\Big].\end{aligned}$$

Thus, the asymptotic variance of $\widehat{\beta}$ is smaller than that of $\widehat{\beta}_1$. That is, $\widehat{\beta}$ is more efficient than $\widehat{\beta}_1$. □

Proof of Theorem 3.4 In model (2.1), $\Lambda_{0k}(t)$ can be rewritten as

$$\Lambda_{0k}(t)=\int_0^t\frac{\mathrm{d}L_k(u)-\beta_0^TG_{1k}(u)\mathrm{d}u}{G_{0k}(u)}.$$

Using uniform strong law of large numbers, we obtain that $\widehat{\Lambda}_{0k}(t,\widehat{\beta})$ converges almost surely to $\Lambda_{0k}(t)$ uniformly in $t\in[0,\tau]$. By Taylor expansion and combining (A.1)-(A.2), we have

$$\begin{aligned}\sqrt{n}(\widehat{\Lambda}_{0k}(t,\widehat{\beta})-\Lambda_{0k}(t)) &= \sqrt{n}(\widehat{\Lambda}_{0k}(t,\widehat{\beta})-\widehat{\Lambda}_{0k}(t,\beta_0))+\sqrt{n}(\widehat{\Lambda}_{0k}(t,\beta_0)-\Lambda_{0k}(t))\\ &= \Big[-\int_0^t\frac{G_{1k}(u)}{G_{0k}(u)}\,\mathrm{d}u\Big]^T\sqrt{n}(\widehat{\beta}-\beta_0)\\ &\quad+\sqrt{n}\Big(\int_0^t\frac{\mathrm{d}\widehat{L}_k(t)-\beta_0^T\widehat{G}_{1k}(u)}{G_{0k}(u)}\\ &\quad-\int_0^t\frac{\widehat{G}_{0k}(u)\mathrm{d}L_k(t)-\beta_0^TG_{1k}(u)\widehat{G}_{0k}(u)\mathrm{d}u}{G_{0k}(u)^2}\Big)+o_p(1)\\ &= n^{-1/2}\sum_{i=1}^n\widehat{\mathcal{E}}_{j_k}\{\Psi_k(t,X_{ikj},\Delta_{ik},Z_{ik})\}+o_p(1).\end{aligned}$$

The finite dimensional normality of $\sqrt{n}(\widehat{\Lambda}_{0k}(t,\widehat{\beta})-\Lambda_{0k}(t))$ can be obtain, by the multivariate central limit theorem. Using the empirical process theory (Pollard, 1990), we can obtain that $\Psi_k(t,X_{ikj},\Delta_{ik},Z_{ik})$ is tight. Thus, $\sqrt{n}(\widehat{\Lambda}_{0k}(t,\widehat{\beta})-\Lambda_{0k}(t))$ converges weakly to a zero-mean Gaussian process with covariance function $\Gamma_k(s,t)$, which can be consistently estimated by $\widehat{\Gamma}_k(s,t)$. $\square$

References

[1] Andersen, P.K. and Gill, R.D., Cox's regression model for counting processes: a large sample study, *The Annals of Statistics*, **10(4)**(1982), 1100–1120.

[2] Cai, J.W. and Schaubel, D.E., Marginal means/rates models for multiple type recurrent event data, *Lifetime Data Analysis*, **10(2)**(2004), 121–138.

[3] Clegg, L.X., Cai, J.W. and Sen, P.K., A marginal mixed baseline hazards model for multivariate failure time data, *Biometrics*, **55(3)**(1999), 805–812.

[4] Huang, Y.J. and Chen, Y.Q., Marginal regression of gaps between recurrent events, *Lifetime Data Analysis*, **9(3)**(2003), 293–303.

[5] Lin, D.Y. and Ying, Z.L., Semiparametric analysis of the additive risk model, *Biometrika*, **81(1)** (1994), 61–71.

[6] Lin, D.Y., Sun, W. and Ying, Z., Nonparametric estimation of the gap time distribution for serial events with censored data, *Biometrika*, **86(1)**(1999), 59–70.

[7] Peña, E., Strawderman, R. and Hollander, M., Nonparametric estimation with recurrent event data, *Journal of the American Statistical Association*, **96(456)**(2001), 1299–1315.

[8] Pollard, D., *Empirical Process: Theory and Applications*, Institute of Mathmatical Statistic: Hayward, 1990.

[9] Prentice, R.L., Williams, B.J. and Peterson, A.V., On the regression analysis of multivariate failure time data, *Biometrika*, **68(2)**(1981), 373–379.

[10] Pepe, M.S. and Cai, J., Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates, *Journal of the American Statistical Association*, **88(423)**(1993), 811–820.

[11] Sun, L.Q., Park, D.H. and Sun, J.G., The additive hazards model for recurrent gap times, *Statistica Sinica*, **16(3)**(2006), 919–932.

[12] Spiekman, C.F. and Lin, D.Y., Marginal regression models for multivariate failure time data, *Journal of the American Statistical Association*, **93(443)**(1998), 1164–1175.

[13] Wang, M.C. and Chang, S.H., Nonparametric estimation of a recurrent survival function, *Journal of the American Statistical Association*, **94(445)**(1999), 146–153.

[14] Wei, L.J., Lin, D.Y. and Weissfeld, L., Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association*, **84(408)**(1989), 1065–1073.

多类型复发事件间隔时间下可加危险率模型

刘吉彩[1] 张日权[1,2] 刘焕彬[3]

([1]华东师范大学金融与统计学院, 上海, 200241; [2]山西大同大学数学系, 大同, 037009)

([3]黄冈师范学院数学与计算机科学学院, 黄冈, 438000)

在许多的生物医学和工程研究中, 多类型复发事件的间隔时间数据是很常见的. 众所周知, 比例危险率模型在一些情况下不能很好拟合生存数据. 本文, 在多类型复发事件的间隔时间数据下, 我们利用可加危险率模型来研究协变量对生存时间的影响程度. 我们采用估计方程方法获得回归系数和基准累积危险率函数估计. 并且, 我们建立了所提估计的渐近分布.

关键词: 可加危险率模型, 间隔时间, 多类型复发事件, 估计方程, 多元生存分析.

学科分类号: O212.07.