

## Study on Asymptotic Property of Additive-Accelerated Mean Regression Model

Huanbin Liu

College of Mathematics and Computer Sciences,  
 Huanggang Normal University, Hubei 438000, China  
 lhb@hgnu.edu.cn

**Keywords:** Recurrent events data; regression model; asymptotic property

**Abstract.** Recurrent event data is a kind of important incomplete data existed in survival analysis, biological medicine research, reliability life test and other practical problems. This paper presents an additive-accelerated mean regression model for multiple type recurrent events data, and gives the estimation methods of unknown parameter and non-parameter function. Specially, the asymptotic properties of parameters estimation are proved.

### Introduction

Recurrent events data is often observed in applied research fields like biostatistics, clinical experiment, and so on. Recurrent events data refers to the reoccurrence time sequence of interested events observed for individuals[1-4]. If only one type of resulted data is concerned, it is referred to as single type recurrent events data. Examples are the recurrent time sequence of acute coronary heart disease and machine faults. If the interested results are of several types, and they may occur several times during an observation period, the data is referred to as multiple type recurrent events data[5-8]. For example, in clinical research, the effect of the infection of pathogens, candida albicans, aspergillosis, and other disease germs on the survival time of kidney-transplantation patients should be studied. In this paper, an additive-accelerated mean regression model for multiple type recurrent events data is presented, and the estimation methods are given. Specially, the asymptotic properties of the estimations are proved.

### The Model

Suppose there are  $n$  individuals to be observed during an observation period, each individual experiences  $k$  different types of recurrent events, and they are mutually independent. Let  $T_{ikj}$  represent the occurrence time of the  $k$ th-type,  $j$ th-time observed event of the  $i$ th individual after the experiment begins, and  $j = 1, 2, \dots, l_k$ , where  $i = 1, 2, \dots, n$ ,  $k = 1, 2, \dots, K$ . Let  $V_{ik}^*(t)$  be the number of recurrence of the  $k$ th type event of the  $i$ th individual at time  $t$ , Definition (1) is given as follows:

$$N_{ik}^*(t) = \sum_{j=1}^n I(T_{ikj} \leq t), \quad (1)$$

where  $I(\cdot)$  is an indicative function. Moreover, suppose the counting process  $N_{ik}^*(t)$  is related to the covariant  $X_{ik}$ ,  $Z_{ik}$ ,  $W_{ik}(t)$  with  $p_1$ ,  $p_2$  and  $p_3$  dimension respectively, and denote as:  $W_{ik}^*(t) = \int_0^t W_{ik}(s)ds$ . The following additive-accelerated mean regression model is suggested to be adopted here.

$$E[N_{ik}^*(t) | X_{ik}, Z_{ik}, W_{ik}(t)] = \mu_0 (te^{\beta_0' X_{ik}}) g(\gamma_0' Z_{ik}) + \alpha_0' W_{ik}^*(t). \quad (2)$$

where  $\beta_0$ ,  $\gamma_0$  and  $\alpha_0'$  are the unknown regression parameter vector of the p1, p2 and p3 dimension respectively, and  $\mu_0(\cdot)$  is the unknown benchmark mean continuous function. Denote  $\tilde{T}_{ikj} = T_{ikj}e^{\beta'X_{ik}}$ , then the counting process is as follows:

$$\tilde{N}_{ik}^*(t; \beta) = \sum_{j=1}^{\infty} I(\tilde{T}_{ikj} \leq t) = \sum_{j=1}^{\infty} I(\tilde{T}_{ikj} \leq te^{-\beta'X_{ik}}) = \tilde{N}_{ik}^*(te^{-\beta'X_{ik}})$$

and model (2) can be expressed as the following form:

$$E[N_{ik}^*(t; \beta_0)] = \mu_0(t)g(\gamma_0'Z_{ik}) + \alpha_0'W_{ik}^*(te^{-\beta_0'X_{ik}}). \quad (3)$$

In many practical applications, individuals are always observed within a limited period, thus  $N_{ik}^*(t)$  can not be observed completely.

### Asymptotic Property and Its Demonstration

To study the asymptotic property of the estimation value of the given model, this chapter assumes the following conditions hold.

(C1) For given  $k$ ,  $\{N_{ik}^*(\cdot), C_{ik}, X_{ik}, Z_{ik}, W_{ik}(\cdot)\}, i=1,2,\dots, n$  is independently and identically distributed, and  $X_{ik}, Z_{ik}$  are linearly independent.

(C2)  $P\{Y_{ik}(\tau, \beta_0) = 1\} > 0$ .

(C3)  $N_{ik}^*(\cdot), X_{ik}, Z_{ik}$  is bounded in  $[0, \tau]$ , and  $W_{ik}(t)$  is a bounded variation function.

(C4)  $g(\cdot)$  is a second order continuous differentiable function, and  $g(\cdot) \geq 0, g(\gamma_0'Z_{ik})$  is locally bounded.

(C5)  $A$  is a nonsingular matrix, and  $A = (A_1, A_2, A_3) \dots$

$$A_1 = E\left[\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta_0) - \bar{x}^*(t)\} Y_{ik}(t; \beta_0) g(\gamma_0'Z_{ik}) X_{ik}' d[\mu_0(t)t]\right],$$

$$A_2 = E\left[\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta_0) - \bar{x}^*(t)\} Y_{ik}(t; \beta_0) g(\gamma_0'Z_{ik}) Z_{ik}' d[\mu_0(t)t]\right],$$

$$A_3 = E\left[\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta_0) - \bar{x}^*(t)\} Y_{ik}(t; \beta_0) e^{-\beta_0'X_{ik}} g W_{ik}'(te^{-\beta_0'X_{ik}}) dt\right].$$

The above are some commonly used regularity conditions. As

$$U_n(\theta) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau X_{ik}^*(t; \beta) - \bar{X}^*(t; \beta, \gamma) [d\tilde{N}_{ik}(t; \beta) - Y_{ik}(t; \beta) e^{-\beta'X_{ik}} \alpha' W_{ik} \alpha' (te^{-\beta'X_{ik}}) dt]$$

to study the asymptotic property of  $\hat{\theta}$ , generally the asymptotic property of  $n^{-\frac{1}{2}}U_n(\theta_0)$  should be proved first.

**Theorem 1.** Under condition (C1)-(C4), the asymptotic property of  $n^{-\frac{1}{2}}U_n(\theta_0)$  obeys normal distribution the mean value of which is zero, and the covariance function is  $\Sigma = E[\varepsilon_i, \varepsilon_i']$ , where

$\varepsilon_i = \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta_0) - \bar{X}^*(t)\} dM_{ik}(t; \theta_0)$ . According to consistent strong law of large numbers, the

consistent estimation of covariant function  $\Sigma$  is  $\hat{\Sigma}$ , where  $\hat{\Sigma} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i(\hat{\theta}) \hat{\varepsilon}_i'(\hat{\theta})$ ,

$$\hat{\varepsilon}_i(\hat{\theta}) = \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \hat{\beta}) - \bar{x}^*(t; \hat{\beta}, \hat{\gamma})\} d\hat{M}_{ik}(t, \hat{\theta}), \quad (4)$$

$$\hat{M}_{ik}(t; \hat{\theta}) = U_n(\theta) = \tilde{N}_{ik}(t; \hat{\beta}) - \int_0^\tau Y_{ik}(s; \hat{\beta}) [g(\hat{\gamma}'Z_{ik}) d\hat{\mu}_0(s; \hat{\theta}) + e^{-\hat{\beta}'X_{ik}} \hat{\alpha}' W_{ik} (se^{-\hat{\beta}'X_{ik}}) ds]. \quad (5)$$

**Proof:** let  $D$  be the compact closed set of  $\theta_0$  neighborhood. Within the closed set of solution  $\hat{\theta}_0$  that contains minimum  $\|U(\theta)\|$ , according to consistent strong law of large numbers,  $n^{-1}U_n(\theta)$  upon  $D$  almost uniformly converges to  $U(\theta)$ .

Let  $dH_{ik}(t; \beta, \alpha) = d\tilde{N}_{ik}(t; \beta) - Y_{ik}(t; \beta)e^{-\beta'X_{ik}}\alpha'W_{ik}(te^{-\beta'X_{ik}})dt$ , and denote

$$U(\theta) = E\left[\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta) - \bar{X}^*(t)\} dH_{ik}(t; \beta, \alpha), S_1(t; \beta, \gamma) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik}) X_{ik}^*(t; \beta), \right.$$

$$S_0(t; \beta, \gamma) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik})$$

Denote  $s_1(t; \beta, \gamma)$ ,  $s_0(t; \beta, \gamma)$  as the limit of  $S_1(t; \beta, \gamma)$ ,  $S_0(t; \beta, \gamma)$ ,  $\bar{X}^*(t) = \frac{s_1(t; \beta, \gamma)}{s_0(t; \beta, \gamma)}$ .

As  $\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau X_{ik}^*(t; \beta) - \bar{X}^*(t; \beta, \gamma) Y_{ik}(t; \beta) g(\gamma'Z_{ik}) d\mu_0(t) = 0$ , then

$$\begin{aligned} U_n(\theta) &= \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta) - \bar{X}^*(t; \beta, \gamma) [d\tilde{N}_{ik}(t; \beta) - Y_{ik}(t; \beta)e^{-\beta'X_{ik}}\alpha'W_{ik}(te^{-\beta'X_{ik}})dt]\} \\ &= \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta) - \bar{X}^*(t; \beta, \gamma)\} dM_{ik}(t; \theta). \end{aligned}$$

Therefore

$$\begin{aligned} n^{-1}U_n(\theta) - U(\theta) &= n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \sum_{k=1}^K X_{ik}^*(t; \beta) dH_{ik}(t; \beta, \alpha) - E\left[\sum_{k=1}^K X_{ik}^*(t; \beta, \alpha) dH_{ik}(t; \beta, \alpha)\right] \right\} \\ &\quad - n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \sum_{k=1}^K \bar{X}^*(t; \beta, \gamma) \{dH_{ik}(t; \beta, \gamma) - E[dH_{ik}(t; \beta, \gamma)]\} \right\} \\ &\quad - n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik}) X_{ik}^*(t; \beta) \frac{E[\sum_{k=1}^K dH_{ik}(t; \beta, \alpha)]}{s_0(t; \beta, \gamma)} \right. \\ &\quad \left. + n^{-1} \sum_{i=1}^n \int_0^\tau E\left[\sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik}) X_{ik}^*(t; \beta) \frac{E[\sum_{k=1}^K dH_{ik}(t; \beta, \alpha)]}{s_0(t; \beta, \gamma)} \right. \right. \\ &\quad \left. \left. + n^{-1} \sum_{i=1}^n \int_0^\tau \sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik}) \frac{S_1(t; \beta, \gamma) E[\sum_{k=1}^K dH_{ik}(t; \beta, \gamma)]}{S_0(t; \beta, \gamma) s_0(t; \beta, \gamma)} \right. \right. \\ &\quad \left. \left. - n^{-1} \sum_{i=1}^n \int_0^\tau E\left[\sum_{k=1}^K Y_{ik}(t; \beta) g(\gamma'Z_{ik}) \right] \frac{S_1(t; \beta, \gamma) E[\sum_{k=1}^K dH_{ik}(t; \beta, \gamma)]}{S_0(t; \beta, \gamma) s_0(t; \beta, \gamma)} \right\} \end{aligned}$$

According to multidimensional central limit theorem:

$$n^{-1}U_n(\theta) - U(\theta) = n^{-1} \sum_{i=1}^n \varepsilon_i(\theta) + O_p n^{-\frac{1}{2}} \quad (6)$$

where  $\varepsilon_i(\theta) = \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta) - \bar{X}^*(t; \beta, \gamma)\} dM_{ik}(t; \theta)$ . Obviously,  $U(\theta_0) = 0$ . Thus,  $n^{-\frac{1}{2}}U_n(\theta_0)$  converges to Gaussian process of zero mean value in terms of distribution, and the covariant of which is  $\Sigma = E\left\{\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta) - \bar{X}^*(t)\} dM_{ik}(t; \theta)^{\otimes 2}\right\}$ .

**Theorem 2.** Under condition (C1)-(C5), the solution  $\hat{\theta}$  of  $U_n(\theta) = 0$  exists uniquely, and is a consistent estimation of  $\theta_0$ . At the same time, the asymptotic property of  $n^{\frac{1}{2}}U_n(\hat{\theta} - \theta_0)$  obeys normal distribution the mean value of which is zero, and the covariance function is  $A^{-1}\Sigma(A^{-1})'$ , where  $A = (A_1, A_2, A_3)$ .

$$\Sigma = E[\{\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \beta_0) - \bar{X}^*(t)\} dM_{ik}(s; \theta_0)\}^{\otimes 2}], a^{\otimes 2} = aa'. \quad (7)$$

Let  $\hat{\lambda}_0(t)$  be the kernel estimation of  $\dot{\mu}_0(t)$ , then we have

$$\hat{\lambda}_0(t) = h^{-1} \int K(\frac{u-t}{h}) d\hat{\mu}_0(u) \quad (8)$$

where  $h$  is the window width,  $K(\cdot)$  is the kernel function, thus the consistent estimation of the asymptotic covariance matrix of  $A^{-1}\Sigma(A^{-1})'$  is  $\hat{A}^{-1}\hat{\Sigma}(\hat{A}^{-1})'$ , where  $\hat{\Sigma} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i(\theta) \hat{\varepsilon}_i'(\theta)$ ,

$$\hat{A} = (\hat{A}_1, \hat{A}_2, \hat{A}_3), \quad \hat{A}_1 = n^{-1} \sum_{i=1}^n [\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \hat{\beta}) - \bar{x}^*(t, \hat{\beta}, \hat{\gamma})\} Y_{ik}(t; \hat{\beta}) g(\gamma_0' Z_{ik}) X_{ik}' d[\hat{\lambda}_0(t)t]],$$

$$\hat{A}_2 = n^{-1} \sum_{i=1}^n [\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \hat{\beta}) - \bar{x}^*(t, \hat{\beta}, \hat{\gamma})\} Y_{ik}(t; \hat{\beta}) \dot{g}(\hat{\gamma}' Z_{ik}) Z_{ik}' d\mu_0(t),$$

$$\hat{A}_3 = n^{-1} \sum_{i=1}^n [\sum_{k=1}^K \int_0^\tau \{X_{ik}^*(t; \hat{\beta}) - \bar{x}^*(t, \hat{\beta}, \hat{\gamma})\} Y_{ik}(t; \hat{\beta}) e^{-\hat{\beta}' X_{ik}} g W_{ik}' (te^{-\beta_0' X_{ik}} dt.$$

The proof process of Theorem 2 is omitted here.

**Theorem 3.** If condition (C1)-(C5) holds, with multiple type recurrent events data,  $\hat{\mu}_0(t, \hat{\theta})$  almost uniformly converges to the benchmark risk function  $\mu_0(t)$  everywhere with  $t \in [0, \tau]$ , and  $n^{\frac{1}{2}}U_n(\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t))$  weakly converges to Gaussian process, the mean value of which is zero and the covariance function is  $\Gamma(s, t) = E[\psi_i(s)\psi_i(t)]$  at  $(s, t)$ .

$$\psi_i(s) = \int_0^\tau \frac{\sum_{k=1}^K dM_{ik}(u; \theta_0)}{s_0(u)} + L(t)' A^{-1} \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(u, \beta_0) - \bar{x}^*(u)\} dM_{ik}(u; \theta_0). \quad (9)$$

where  $L(t) = -(L_1(t)', L_2(t)', L_3(t)')$ ,

$$L_1(t) = \int_0^\tau \frac{E[\sum_{k=1}^K Y_{ik}(s; \beta_0) g(\gamma_0' Z_{ik}) X_{ik}]}{s_0(s)} d\{\dot{\mu}_0(s)s\}, \quad L_2(t) = \int_0^\tau \frac{E[\sum_{k=1}^K Y_{ik}(s; \beta_0) \dot{g}(\gamma_0' Z_{ik}) X_{ik}]}{s_0(s)} d\{\mu_0(s),$$

$$L_3(t) = \int_0^\tau \frac{E[\sum_{k=1}^K Y_{ik}(s; \beta_0)] e^{-\beta_0' X_{ik}} W_{ik}(se^{-\beta_0' X_{ik}})}{s_0(s)} ds \{X_{ik}^*(t; \hat{\beta})\}, \quad s_0(t) = E[\sum_{k=1}^K Y_{ik}(t; \beta_0)] g(\gamma_0' Z_{ik}).$$

The estimation of  $s_0(t)$ ,  $L(t)$  and  $L_i(t)$ ,  $i = 1, 2, 3$  are respectively

$$\hat{L}(t) = -(\hat{L}_1(t)', \hat{L}_2(t)', \hat{L}_3(t)')', \quad \hat{L}_1(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\sum_{k=1}^K Y_{ik}(s; \hat{\beta}) g(\hat{\gamma}' Z_{ik}) X_{ik}}{\hat{S}_0(s)} d\{\hat{\lambda}_0(s)s\},$$

$$\hat{L}_2(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\sum_{k=1}^K Y_{ik}(s; \hat{\beta}) \dot{g}(\gamma' Z_{ik}) Z_{ik}}{\hat{S}_0(s)} d\{\hat{\mu}_0(s)\}, \quad \hat{S}_0(t) = \frac{1}{n} \sum_{i=1}^n \left[ \sum_{k=1}^K Y_{ik}(t; \hat{\beta}) \right] g(\gamma' Z_{ik}),$$

$$\hat{L}_3(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\sum_{k=1}^K Y_{ik}(s; \hat{\beta}) e^{-\hat{\beta}_0' X_{ik}} W_{ik}(s e^{-\hat{\beta}_0' X_{ik}})}{\hat{S}_0(s)} ds.$$

Thus the uniform estimation of  $\Gamma(s, t)$  is  $\hat{\Gamma}(s, t) = n^{-1} \sum_{i=1}^n \hat{\psi}_i(s) \hat{\psi}_i(t)$ , where

$$\hat{\psi}_i(s) = \int_0^\tau \frac{\sum_{k=1}^K d\hat{M}_{ik}(u; \hat{\theta})}{\hat{S}_0(u)} + \hat{L}(t)' \hat{A}^{-1} \sum_{k=1}^K \int_0^\tau \{X_{ik}^*(u; \hat{\theta}) - \bar{X}(u; \hat{\beta}, \hat{\gamma})\} d\hat{M}_{ik}(u; \hat{\theta}).$$

When estimating the asymptotic function of  $n^{\frac{1}{2}} U_n(\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t))$ , the hazard ratio  $\lambda_0(t)$  also should be estimated, however, the tail of the benchmark hazard ratio tends to be unstable in the presence of censored data, in this case, the asymptotic function of  $n^{\frac{1}{2}} U_n(\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t))$  is difficult to be estimated directly, so we adopt the method proposed by Lin, Wei and Ying(1998), that's, the distribution of  $n^{\frac{1}{2}} U_n(\hat{\mu}_0(t, \hat{\theta}) - \mu_0(t))$  can be represented by estimating  $\hat{V}(t)$ .

$$\hat{V}(t) = n^{\frac{1}{2}} \{ \hat{\mu}_0(t, \hat{\theta}) - \hat{\mu}_0(t, \hat{\theta}^*) \} + n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \frac{d\hat{M}_{ik}(s, \hat{\theta})}{\hat{S}_0(s)} G_{ik}$$

The proof process of Theorem 3 is omitted here.

## Conclusions

In this paper, an additive-accelerated mean regression model for multiple type recurrent events is presented, and the estimation methods of unknown parameter and non-parameter function are given. In addition, the asymptotic properties of parameters estimation are proved under the case of large-scale samples.

## Acknowledgments

This work is supported by the Natural Science Foundation of Hubei Province, China (No. 2011CDB167), the Major Research Program of Hubei Provincial Department of Education, China (No. Z20092701), the Ph.D. Fund of Huanggang Normal University to H. B. Liu, and the Innovative Group Project of Hubei Provincial Department of Education (No. 03BA85).

## References

- [1] U. Uzunogullari, J.L. Wang, *Biometrika* 79 (1992) 297-310.
- [2] T.L. Lai, Z. Ying, *Ann. Statist.* 19 (1991) 531-556.
- [3] K.Y. Liang, S.G. Self, Y. Chang, J. Roy. *Statist. Soc. Ser. B* 55 (1993) 441-453.
- [4] Y.D. Lin, Z. Ying, *Biometrika* 81 (1994) 61-71.
- [5] C.A. McGilchrist, *Biometrics* 49 (1993) 221-225.
- [6] H. Aslanidou, D.K. Dey, D. Sinha, *Can. J. Statist.* 26 (1998) 33-48.
- [7] I.D. Ha, Y. Lee, J.K. Song, *Biometrika* 88 (2001) 233-243.
- [8] Y. Li, X. Lin, *Biometrika* 87 (2000) 849-866.