Statistica Sinica **13**(2003), 1237-1257

ADDITIVE HAZARDS REGRESSION WITH MISSING CENSORING INFORMATION

Xian Zhou and Liuquan Sun

Hong Kong Polytechnic University and Academia Sinica, Beijing

Abstract: In this article we study estimation in the *additive* hazards regression model with missing censoring indicators. We develop simple procedures to obtain consistent and efficient estimators for the regression parameters as well as the cumulative baseline hazard function, and derive their asymptotic properties. The estimator of the regression parameters is shown to be asymptotically normally distributed, while the estimator of the cumulative baseline hazard function converges to a Gaussian process. We address both the situations where the mechanism for missingness of the censoring indicators is independent of any other factors, and those in which the missingness may depend on the covariates. Monte Carlo studies are also conducted to evaluate the performance of the estimators.

Key words and phrases: Additive risk, censoring, estimating equation, incomplete data, Markov process, missing at random.

1. Introduction

When analyzing censored survival data the censoring information may be missing for some individuals (e.g., autopsy for some subjects may not be carried out to save expense, or medical records may be missing). Survival models with missing censoring indicators have been studied by several authors, notably Lo (1991), Gijbels, Lin and Ying (1993), McKeague and Subramanian (1998), van der Laan and McKeague (1998) and Subramanian (2000), among others.

In this article we consider the regression analysis of failure times with censoring indicators missing at random. Specifically, let T denote the failure time, $Z(\cdot)$ be a time-varying covariate vector, and C be a censoring time that is conditionally independent of T given $Z(\cdot)$. Data are available on $X = T \wedge C$ and $Z(t)$ $(0 \leq t \leq X)$, but the censoring indicator $\delta = I(T \leq C)$ may be missing. When the mechanism for missingness of the censoring indicators is independent of everything else, it is referred to as *missing completely at random* (MCAR), while the more general case of missing at random is known as MAR, see Little and Rubin (1987).

In the absence of covariates and under MCAR, Dinse (1982) obtained a nonparametric maximum likelihood estimator (NPMLE) of the survival function using the EM algorithm. Lo (1991) proved that there are infinitely many NPMLES, some of which may be inconsistent. In addition, ad hoc estimators have been proposed by Lo (1991), Gijbels, Lin and Ying (1993) and McKeague and Subramanian (1998). Under the more general MAR scenario, van der Laan and McKeague (1998) proposed a sieved nonparametric maximum likelihood estimator, and showed that it is asymptotically efficient.

When covariates are present, Gijbels, Lin and Ying (1993) initiated research on estimation for the Cox model under MCAR; McKeague and Subramanian (1998) provided an alternative approach to estimation; Subramanian (2000) considered estimation under proportionality of conditional hazards; Goetghebeur and Ryan (1995) analyzed competing risks survival data with proportional hazards regression models under MAR, and presented consistent and asymptotically normal estimators of the regression parameters and related score tests. To date, however, the estimation for the *additive hazards* (AH) model (cf. Lin and Ying (1994)) with missing censoring indicators appears to have not been addressed.

The AH model specifies that the hazard function associated with a set of covariates is the sum of the baseline hazard function and the regression function of the covariates. It has been found to be more plausible than the Cox model for many applications, see Aalen (1980), Cox and Oakes (1984), Breslow and Day (1987), Lin and Ying (1994) and McKeague and Sasieni (1994). The main objective of this article is to study the estimation of the regression parameters as well as the baseline hazard function for the AH model with missing censoring indicators. We first derive consistent estimators and their asymptotic properties under MCAR. As the cause of failure may well be influenced by covariates (e.g., age, gender, or treatment method), we also study an extension in which the missingness depends on the covariates.

The rest of the paper is organized as follows. The estimation of the regression parameters under MCAR is given in Section 2. In Section 3, we provide an estimator of the cumulative baseline hazard function as well as its asymptotic properties under MCAR. The extension to the case of the missingness dependent on the covariates is considered in Section 4. Some Monte Carlo studies on the proposed estimators are presented in Section 5, followed by some concluding remarks in Section 6. Finally, detailed proofs are provided Section 7.

2. Estimation of Regression Parameters under MCAR

We first briefly review the AH model studied by Lin and Ying (1994), and describe their proposed estimator. In the AH model, the hazard function for a failure time T is assumed to be of the form

$$\lambda(t \mid Z) = \lambda_0(t) + \beta_0' Z(t), \tag{2.1}$$

where $Z(t)$ is a p-vector of possibly time-varying covariates, $\lambda_0(t)$ is an unspecified baseline hazard function, β_0 is a p-vector of unknown regression parameters, and v' denotes the transpose of a vector or matrix v. Also let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ be the cumulative baseline hazard function.

In the case where the data are fully observed, they consist of independent triplets $(X_i, \delta_i, Z_i(t); 0 \le t \le X_i), i = 1, \cdots, n$. Lin and Ying (1994) introduced a pseudoscore function

$$U_0(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \{Z_i(t) - \bar{Z}(t)\}\{dN_i^u(t) - Y_i(t)\beta' Z_i(t)dt\} \tag{2.2}$$

as an estimating function for the parameter vector β_0, where $N_i^u(t) = I(X_i \le t, \delta_i = 1)$, $Y_i(t) = I(X_i \ge t)$, and $\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i(t)/\sum_{i=1}^n Y_i(t)$. The resulting estimator, which takes an explicit form, is consistent and asymptotically normal with an easily estimated covariance matrix. In addition, the cumulative baseline hazard $\Lambda_0(t)$ can be consistently estimated by

$$\tilde{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i^u(s) - Y_i(s)\tilde{\beta}' Z_i(s)ds\}}{\sum_{i=1}^n Y_i(s)}, \tag{2.3}$$

which converges weakly to a Gaussian process, where $\tilde{\beta}$ is an estimate of β_0.

In the MCAR case, the observed data consist of n independent and identically distributed vectors $(X_i, \xi_i, \sigma_i, Z_i(t),\ 0 \le t \le X_i)$ $(i = 1, \ldots, n)$, where ξ_i is the indicator that δ_i is not missing and $\sigma_i = \xi_i \delta_i$. It is assumed that ξ_i is independent of $(T_i, C_i, Z_i(\cdot))$, and $\rho = Pr(\xi_i = 1) > 0$ under MCAR. A naive method for estimating β_0 is to simply ignore the missing data and apply the pseudoscore function U_0 to the complete data only. Such a procedure (called the complete case estimator) is clearly inefficient if there is a significant amount of missing data.

Let $H_{jk}(t) = P(X_i \le t, \xi_i = j, \sigma_i = k)$ for $(j,k) \in \{(1,1),(1,0),(0,0)\}$, and define $d\Lambda_{jk}(t) = dH_{jk}(t)/\bar{H}_-(t)$, where $\bar{H}(t) = 1 - \sum_{(j,k)\in\Delta} H_{jk}(t)$, and $\bar{H}_-(t) = \bar{H}(t-)$. Then the cumulative hazard function of T (denoted by Λ_T) can be expressed as (McKeague and Subramanian (1998))

$$\Lambda_T(t \mid Z) = \Lambda_{11}(t \mid Z) + \pi(t \mid Z)\Lambda_{00}(t \mid Z), \tag{2.4}$$

where $\pi = \Lambda_{11}/(\Lambda_{11} + \Lambda_{10})$. Define the following pseudoscore functions (cf. Lin and Ying (1994)):

$$U_{11}(\beta, t) = \sum_{i=1}^{n} \int_0^t \{Z_i(s) - \bar{Z}(s)\}\xi_i\{dN_i^u(s) - Y_i(s)\beta' Z_i(s)ds\},$$

$$U_{10}(\beta,t)=\sum_{i=1}^{n}\int_0^t\{Z_i(s)-\bar{Z}(s)\}\xi_i dN_i^c(s), \tag{2.5}$$
$$U_{00}(\beta,t)=\sum_{i=1}^{n}\int_0^t\{Z_i(s)-\bar{Z}(s)\}(1-\xi_i)\{dN_i(s)-Y_i(s)\beta' Z_i(s)ds\},$$

where $N_i(t)=I(X_i\le t)$ and $N_i^c(t)=(1-\delta_i)N_i(t)$. In view of (2.4), and following the idea of McKeague and Subramanian (1998) for the Cox model, we propose the following estimating function:

$$U(\beta,t)=U_{11}(\beta,t)+P(\beta,t)U_{00}(\beta,t), \tag{2.6}$$

where $P(\beta,t)=\mathrm{diag}(U_{11}(\beta,t))[\mathrm{diag}(U_{11}(\beta,t))+\mathrm{diag}(U_{10}(\beta,t))]^{-1}$, and $\mathrm{diag}(v)$ is the diagonal matrix with diagonal vector v. Let $0<\tau<\infty$ be a fixed quantity. Our proposed estimator $\hat{\beta}$ is a solution to the estimating equation $U(\beta,\tau)=0$. It can be shown that there exists a neighborhood of β_0 within which, with probability approaching 1 as $n\to\infty$, the root $\hat{\beta}$ of $U(\beta,\tau)=0$ is uniquely defined (see the proof of Theorem 1). Let $\bar{z}(t)=E[Y_i(t)Z_i(t)]/E[Y_i(t)]$. To establish the asymptotic properties of $\hat{\beta}$, we need the following regularity conditions:

R1. $\Lambda_0(\tau)<\infty$;
R2. $Pr(Y_i(\tau)=1)>0$;
R3. $E[\sup_{0\le t\le\tau}\|Z_i(t)\|^2]<\infty$;
R4. $A=E[\int_0^\tau(Z_i(t)-\bar{z}(t))^{\otimes 2}Y_i(t)dt]$ is nonsingular, where $a^{\otimes 2}=aa'$ for any vector a.

The asymptotic properties of $\hat{\beta}$ are given in the following theorem.

Theorem 1. *Under R1−R4, the estimator $\hat{\beta}$ of β_0 is consistent and $n^{1/2}(\hat{\beta}-\beta_0)$ converges in distribution to a zero-mean normal random vector with variance*

$$V=A^{-1}\Sigma A^{-1}+(\rho^{-1}-1)A^{-1}E\{N_1^{CZ}(\tau)-B(\tau)N_1^Z(\tau)\}^{\otimes 2}A^{-1}, \tag{2.7}$$

where $\Sigma=E[\int_0^\tau\{Z_i(s)-\bar{z}(s))\}^{\otimes 2}dN_i^u(s)]$,

$$N_i^{CZ}(t)=\int_0^t\{Z_i(s)-\bar{z}(s)\}dN_i^c(s), \tag{2.8}$$
$$N_i^{Z}(t)=\int_0^t\{Z_i(s)-\bar{z}(s)\}\{dN_i(s)-Y_i(s)\beta_0' Z_i(s)ds\}, \tag{2.9}$$

and $B(t)=\mathrm{diag}\{E[N_{1j}^{CZ}(t)]/E[N_{1j}^{Z}(t)],\ j=1,\dots,p\}$.

Note that the first term in (2.7) is the asymptotic variance of the estimator in Lin and Ying (1994) based on the full data ($\rho=1$), and the second term represents the effect of the missing censoring indicators.

Let $\hat{\rho} = n^{-1}\sum_{i=1}^{n}\xi_i$. By plugging $\hat{\beta}$ and $\hat{\rho}$ into the corresponding empirical estimator in place of the unknown β_0 and ρ, and replacing the (unobserved) processes N_i^u and N_i^c by $\hat{\rho}^{-1}\xi_i N_i^u$ and $\hat{\rho}^{-1}\xi_i N_i^c$ (which are observable), respectively,

$$\hat{V} = \hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1} + (\hat{\rho}^{-1} - 1)\hat{A}^{-1}\hat{W}\hat{A}^{-1} \tag{2.10}$$

is a consistent estimator of V, where

$$\begin{aligned}
\hat{A} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s) - \bar{Z}(s)\}^{\otimes 2} Y_i(s)ds,\\
\hat{\Sigma} &= \hat{\rho}^{-1}n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s) - \bar{Z}(s)\}^{\otimes 2}\xi_i dN_i^u(s),\\
\hat{W} &= n^{-1}\sum_{i=1}^{n}\{\hat{N}_i^{CZ}(\tau) - \hat{B}(\tau)\hat{N}_i^{Z}\}^{\otimes 2},\\
\hat{N}_i^{CZ}(t) &= \hat{\rho}^{-1}\int_0^t \{Z_i(s) - \bar{Z}(s)\}\xi_i dN_i^c(s),\\
\hat{N}_i^{Z}(t) &= \int_0^t \{Z_i(s) - \bar{Z}(s)\}\{dN_i(s) - Y_i(s)\hat{\beta}' Z_i(s)ds\},\\
\hat{B}(t) &= \mathrm{diag}\Big\{n^{-1}\sum_{i=1}^{n}\hat{N}_{ij}^{CZ}(t)/n^{-1}\sum_{i=1}^{n}\hat{N}_{ij}^{Z}(t),\ j = 1,\ldots,p\Big\}.
\end{aligned}$$

3. Estimation of Cumulative Baseline Hazard Function under MCAR

The cumulative baseline hazard function Λ_0 can be expressed in the same form as the basic equation (2.4) for Λ_T,

$$\Lambda_0(t) = \Lambda_{11}^0(t) + \pi^0(t)\Lambda_{00}^0(t), \tag{3.1}$$

where

$$\begin{aligned}
\Lambda_{11}^0(t) &= \rho\int_0^t \frac{E[dN_1^u(s) - Y_1(s)\beta_0' Z_1(s)ds]}{E[Y_1(s)]},\\
\Lambda_{10}^0(t) &= \rho\int_0^t \frac{dE[N_1^c(s)]}{E[Y_1(s)]},\\
\Lambda_{00}^0(t) &= (1-\rho)\int_0^t \frac{E[dN_1(s) - Y_1(s)\beta_0' Z_1(s)ds]}{E[Y_1(s)]},\\
\pi^0(t) &= \frac{\Lambda_{11}^0(t)}{\Lambda_{11}^0(t) + \Lambda_{10}^0(t)}.
\end{aligned}$$

This leads to the following estimator of the cumulative baseline hazard function:

$$\hat{\Lambda}_0(t) = \hat{\Lambda}_{11}^0(t) + \hat{\pi}^0(t)\hat{\Lambda}_{00}^0(t), \tag{3.2}$$

where

$$\begin{aligned}
\hat{\Lambda}_{11}^0(t) &= \int_0^t \frac{\sum_{i=1}^n \xi_i[dN_i^u(s) - Y_i(s)\hat{\beta}' Z_i(s)ds]}{\sum_{i=1}^n Y_i(s)}, \\
\hat{\Lambda}_{10}^0(t) &= \int_0^t \frac{\sum_{i=1}^n \xi_i dN_i^c(s)}{\sum_{i=1}^n Y_i(s)}, \\
\hat{\Lambda}_{00}^0(t) &= \int_0^t \frac{\sum_{i=1}^n (1-\xi_i)[dN_i(s) - Y_i(s)\hat{\beta}' Z_i(s)ds]}{\sum_{i=1}^n Y_i(s)}, \\
\hat{\pi}^0(t) &= \frac{\hat{\Lambda}_{11}^0(t)}{\hat{\Lambda}_{11}^0(t) + \hat{\Lambda}_{10}^0(t)},
\end{aligned}$$

where the last is defined to be zero when the denominator vanishes. The estimator $\hat{\Lambda}_0(t)$ reduces to the Lin and Ying estimator $\tilde{\Lambda}_0(t)$ (see (2.3)) when there are no missing censoring indicators. The asymptotic properties of the estimator $\hat{\Lambda}_0(t)$ are given in the next theorem.

Theorem 2. *Under the assumptions of Theorem* 1, *we have* (i) $\sup_{0\le t\le\tau} |\hat{\Lambda}_0(t) - \Lambda_0(t)| \xrightarrow{p} 0$; (ii) $n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\}$ *converges weakly on* $[0,\tau]$ *to a zero-mean Gaussian process whose covariance function at* (t,s) $(t \le s)$ *is*

$$\begin{aligned}
G(t,s) = {} & C_1(t,s)\int_0^t \frac{[\lambda_0(u) + \beta_0'\bar{z}(u)]du}{E[Y_1(u)]} + C_2(t,s)\int_0^t \frac{dE[N_1^c(u)]}{(E[Y_1(u)])^2} \\
& + C_3(t,s)\int_0^t\int_u^t \frac{dE[N_1^c(v)]d\Lambda_0(u)}{E[Y_1(v)]E[Y_1(u)]} + C_3(s,t)\int_0^t\int_u^s \frac{dE[N_1^c(v)]d\Lambda_0(u)}{E[Y_1(v)]E[Y_1(u)]} \\
& + C_4(t,s)\int_0^t \frac{[\Lambda_0(t) + \Lambda_0(s) - 2\Lambda_0(u)]\lambda_0(u)du}{E[Y_1(u)]} \\
& + C_5(t,s)[d(t)'A^{-1}\Sigma A^{-1}d(s) - d(t)'A^{-1}D(s) - d(s)'A^{-1}D(t)], \qquad (3.3)
\end{aligned}$$

where

$$\begin{aligned}
C_1(t,s) = {} & \{1+\rho D_{00}(t)D_{10}(t) + (1-\rho)D_{11}(t)\}\{1+\rho D_{00}(s)D_{10}(s) + (1-\rho)D_{11}(s)\} \\
& + \rho(1-\rho)\{1 + D_{00}(t)D_{10}(t) - D_{11}(t)\}\{1 + D_{00}(s)D_{10}(s) - D_{11}(s)\}, \\
C_2(t,s) = {} & D_{11}(t)D_{11}(s)\{1-\rho-\rho D_{00}(t)\}\{1-\rho-\rho D_{00}(s)\} \\
& + \rho(1-\rho)D_{11}(t)D_{11}(s)\{1 + D_{00}(t)\}\{1 + D_{00}(s)\}, \\
C_3(t,s) = {} & -\rho(1-\rho)D_{11}(t)(1 + D_{00}(t))\{1 + D_{00}(s)D_{10}(s) - D_{11}(s)\}, \\
C_4(t,s) = {} & \rho(1-\rho)\{1 + D_{00}(t)D_{10}(t) - D_{11}(t)\}\{1 + D_{00}(s)D_{10}(s) - D_{11}(s)\}, \\
C_5(t,s) = {} & \{1+\rho D_{00}(t)D_{10}(t) + (1-\rho)D_{11}(t)\}\{1+\rho D_{00}(s)D_{10}(s) + (1-\rho)D_{11}(s)\},
\end{aligned}$$

$$D_{kl}(t,s) = \frac{\Lambda_{kl}^0(t)}{\Lambda_{11}^0(t) + \Lambda_{10}^0(t)}, \quad k,l = 0,1, \qquad (3.4)$$

$$d(t) = \int_0^t \bar{z}(s)ds, \tag{3.5}$$
$$D(t) = \int_0^t \left\{ \frac{E[Y_1(s)Z_1(s)^{\otimes 2}]}{E[Y_1(s)]} - \bar{z}(s)^{\otimes 2} \right\} \beta_0 du.$$

The covariance in (3.3) can be consistently estimated by plugging $\hat{\beta}$ and $\hat{\rho}$ into the corresponding empirical estimator in place of the unknown β_0 and ρ, and replacing the (unobserved) processes N_i^u and N_i^c by $\hat{\rho}^{-1}\xi_i N_i^u$ and $\hat{\rho}^{-1}\xi_i N_i^c$, respectively. For an individual with a given covariate vector $z_0(\cdot)$, the corresponding estimator of the survival function $S(t, z_0)$ is

$$\hat{S}(t, z_0) = \exp\{-\hat{\Lambda}_0(t) - \int_0^t \hat{\beta}' z_0(u)du\}. \tag{3.6}$$

By the functional delta-method and Theorem 2, we can get the asymptotic properties of $\hat{S}(t, z_0)$ and these can be applied to construct confidence bands for $S(t, z_0)$.

Remark 1. As shown in (3.2), the estimator $\hat{\Lambda}_0(t)$ may not always be monotone in t. However the modification suggested by Lin and Ying (1994), i.e., $\tilde{\Lambda}_0^*(t) = \max_{s \le t} \hat{\Lambda}_0(t)$, can ensure the monotonicity of $\tilde{\Lambda}_0^*(t)$ and, under appropriate regularity conditions, $\tilde{\Lambda}_0^*(t)$ and $\hat{\Lambda}_0(t)$ are asymptotically equivalent in the sense that $\tilde{\Lambda}_0^*(t) - \hat{\Lambda}_0(t) = o_p(n^{-1/2})$.

4. Missingness Dependent on Covariates

In this section we consider an extension to allow the missing machanism to depend on the covariates, but under the restriction that the covariates are time invariant. We assume that, conditional on the covariate Z, the missingness indicator ξ is independent of T and C.

It is typical in this case to specify a parametric model for the missing data mechanism (Rubin (1976)), that is,

$$P(\xi_i = 1 \mid T_i, C_i, Z_i) = \phi(Z_i, \alpha), \tag{4.1}$$

where ϕ is a known function and α is an unknown parameter vector distinct from β. Often we can specify ϕ as a logistic regression function since ξ_i is binary (cf. Ibrahim, Lipsitz and Chen (1999) or Lipsitz, Ibrahim and Zhao (1999)). Assume ϕ is twice differentiable with respect to α. Because Z_i is always observed, we can get the consistent and asymptotically normal maximum likelihood estimate $\hat{\alpha}$ of α based on $\{Z_i\}$.

Let $\rho_i = \phi(Z_i, \alpha_0)$, $\hat{\rho}_i = \phi(Z_i, \hat{\alpha})$, where α_0 is the true value of α. Using the weighted estimating equation procedure (cf. Lipsitz, Ibrahim and Zhao (1999)), we propose the following estimating function for β_0 under the missingness dependent on the covariates:

$$U^*(\beta,t) = U^*_{11}(\beta,t) + P^*(\beta,t)U^*_{00}(\beta,t), \tag{4.2}$$

where

$$\begin{aligned}
U^*_{11}(\beta,t) &= \sum_{i=1}^n \int_0^t \{Z_i - \bar Z(s)\}\frac{\xi_i}{\hat\rho_i}\{dN_i^u(s) - Y_i(s)\beta' Z_i ds\},\\
U^*_{10}(\beta,t) &= \sum_{i=1}^n \int_0^t \{Z_i - \bar Z(s)\}\frac{\xi_i}{\hat\rho_i} dN_i^c(s),\\
U^*_{00}(\beta,t) &= \sum_{i=1}^n \int_0^t \{Z_i - \bar Z(s)\}(1 - \frac{\xi_i}{\hat\rho_i})\{dN_i(s) - Y_i(s)\beta' Z_i ds\},\\
P^*(\beta,t) &= \mathrm{diag}(U^*_{11}(\beta,t))[\mathrm{diag}(U^*_{11}(\beta,t)) + \mathrm{diag}(U^*_{10}(\beta,t))]^{-1}.
\end{aligned}$$

Our estimator $\hat\beta^*$ is a solution to the estimating equation $U^*(\beta,\tau) = 0$, which reduces to $\hat\beta$ under MCAR. It can be checked that there exists a unique root $\hat\beta^*$ of $U(\beta,\tau) = 0$ in a neighborhood of β_0 with probability approaching 1 as $n \to \infty$. The asymptotic properties of $\hat\beta^*$ are established in the following theorem.

Theorem 3. *Assume that ϕ is twice differentiable with respect to α, and* R1−R4 *hold. Then the estimator $\hat\beta^*$ of β_0 is consistent and $n^{1/2}(\hat\beta^* - \beta_0)$ converges in distribution to a zero-mean normal random vector with variance $A^{-1}(\Sigma+\Phi)A^{-1}$, where*

$$\begin{aligned}
\Phi &= E\Big\{\rho_1^{-1}(1-\rho_1)^{-1}\Big((1-\rho_1)[N_1^{CZ}(\tau) - B(\tau)N_1^Z(\tau)] - \Omega\Gamma^{-1}\frac{\partial\phi(Z_1,\alpha_0)}{\partial\alpha}\Big)^{\otimes 2}\Big\},\\
\Omega &= E\Big(\rho_1 N_1^{CZ}(\tau)\frac{\partial\phi(Z_1,\alpha_0)}{\partial\alpha'}\Big),\\
\Gamma &= -E\Big[\rho_1\frac{\partial^2\log\phi(Z_1,\alpha_0)}{\partial\alpha\partial\alpha'} + (1-\rho_1)\frac{\partial^2\log(1-\phi(Z_1,\alpha_0))}{\partial\alpha\partial\alpha'}\Big].
\end{aligned}$$

The asymptotic variance matrix $A^{-1}(\Sigma+\Phi)A^{-1}$ can be consistently estimated by $\hat A^{-1}(\hat\Sigma^* + \hat\Phi)\hat A^{-1}$, where

$$\begin{aligned}
\hat\Sigma^* &= n^{-1}\sum_{i=1}^n \int_0^\tau \{Z_i - \bar Z(s)\}^{\otimes 2}\frac{\xi_i}{\hat\rho_i} dN_i^u(s),\\
\hat\Phi &= n^{-1}\sum_{i=1}^n \hat\rho_i^{-1}(1-\hat\rho_i)^{-1}\Big[(1-\hat\rho_i)[\tilde N_i^{CZ}(\tau) - \tilde B(\tau)\hat N_i^Z(\tau)] - \hat\Omega\hat\Gamma^{-1}\frac{\partial\phi(Z_i,\hat\alpha)}{\partial\alpha}\Big]^{\otimes 2},\\
\tilde N_i^{CZ}(t) &= \int_0^t \{Z_i - \bar Z(s)\}\frac{\xi_i}{\hat\rho_i} dN_i^c(s),
\end{aligned}$$

$$\tilde{B}(t) = \mathrm{diag}\Big\{ n^{-1}\sum_{i=1}^{n} \tilde{N}_{ij}^{CZ}(t)/n^{-1}\sum_{i=1}^{n} \hat{N}_{ij}^{Z}(t), j = 1, \cdots, p \Big\},$$

$$\hat{\Omega} = n^{-1}\sum_{i=1}^{n} \hat{\rho}_i \tilde{N}^{CZ}(\tau)\frac{\partial \phi(Z_i, \hat{\alpha})}{\partial \alpha'},$$

$$\hat{\Gamma} = -n^{-1}\sum_{i=1}^{n} \Big[\xi_i \frac{\partial^2 \log \phi(Z_i, \hat{\alpha})}{\partial \alpha \partial \alpha'} + (1-\xi_i)\frac{\partial^2 \log(1-\phi(Z_i, \hat{\alpha}))}{\partial \alpha \partial \alpha'} \Big].$$

$\hat{A}$ and $\hat{N}_i^Z(t)$ are defined following (2.10), and $\bar{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i/\sum_{i=1}^n Y_i(t)$ (which is still time dependent although Z_i are time invariant in this section).

Similarly, we can estimate the cumulative baseline hazard function Λ_0 by

$$\hat{\Lambda}_0^*(t) = \hat{\Lambda}_{11}^*(t) + \hat{\pi}^*(t)\hat{\Lambda}_{00}^*(t), \tag{4.3}$$

where

$$\hat{\Lambda}_{11}^*(t) = \int_0^t \frac{\sum_{i=1}^n \frac{\xi_i}{\hat{\rho}_i}[dN_i^u(s) - Y_i(s)\hat{\beta}' Z_i ds]}{\sum_{i=1}^n Y_i(s)}, \tag{4.4}$$

$$\hat{\Lambda}_{10}^*(t) = \int_0^t \frac{\sum_{i=1}^n \frac{\xi_i}{\hat{\rho}_i} dN_i^c(s)}{\sum_{i=1}^n Y_i(s)}, \tag{4.5}$$

$$\hat{\Lambda}_{00}^*(t) = \int_0^t \frac{\sum_{i=1}^n (1 - \frac{\xi_i}{\hat{\rho}_i})[dN_i(s) - Y_i(s)\hat{\beta}' Z_i ds]}{\sum_{i=1}^n Y_i(s)}, \tag{4.6}$$

$$\hat{\pi}^*(t) = \frac{\hat{\Lambda}_{11}^*(t)}{\hat{\Lambda}_{11}^*(t) + \hat{\Lambda}_{10}^*(t)}. \tag{4.7}$$

The estimator $\hat{\Lambda}_0^*(t)$ reduces to $\hat{\Lambda}_0(t)$ under MCAR. As in Theorem 2, we can prove that $\hat{\Lambda}_0^*(t)$ is a consistent estimate of $\Lambda_0(t)$, and $n^{1/2}\{\hat{\Lambda}_0^*(t) - \Lambda_0(t)\}$ converges weakly on $[0, \tau]$ to a zero-mean Gaussian process whose covariance function is more complicated than (3.3). We omit the details.

Remark 2. If the binary regression function ϕ is unknown, we estimate it by some nonparametric method, such as kernel smoothing, wavelet method, spline approximation, or local polynomial modelling, etc. As shown in McKeague and Subramanian (1998), it would be a non-trivial problem for the nonparametric method because the 'curse-of-dimensionality' implies that ϕ is difficult to estimate for high-dimensional covariates. But if there exists a consistent estimator of ϕ, then we can estimate β_0 as in (4.2).

5. Simulation Studies

We carried out a small Monte Carlo study to compare the performance of the proposed estimators with those of the complete case estimators of β_0 and Λ_0 under the MCAR model. The case when missingness is dependent on the covariates is similar, though not reported here. The underlying AH model was taken to be $\lambda(t \mid Z) = 1 + \beta_0' Z$ for $\beta_0 = 0, 0.5, 1$, where Z is uniformly distributed on $(0, 6)$. The censoring was exponential with the parameter adjusted to give prescribed censoring rates. In each case the mean square errors (MSE) of the various estimators of β_0 were computed from 1,000 simulated samples of size $n = 200$ each. The results are shown in Tables 1−2. The "full data" estimator is also included for comparison. The results in Tables 1−2 are classified according to the value of ρ (0.8 or 0.5). We have also tested selected simulations with 2,000, 5,000 or 10,000 replications, the results are very similar to those in Tables 1 and 2. Figures 1−2 show the plots of the various estimators of Λ_0 with $\beta_0 = 0.5$ and a single simulation in each case.

Table 1. Simulation results with $\rho = 0.8$.

		Full data		Proposed		Complete case	
$P(\delta = 0)$	β_0	Mean	MSE	Mean	MSE	Mean	MSE
	0.2	0.2011	0.0053	0.2026	0.0059	0.2048	0.0080
25%	0.5	0.5026	0.0133	0.5042	0.0149	0.5099	0.0198
	1.0	1.0107	0.0338	1.0140	0.0372	1.0230	0.0488
	0.2	0.2017	0.0174	0.2031	0.0193	0.2095	0.0239
75%	0.5	0.5096	0.0391	0.5104	0.0446	0.5128	0.0570
	1.0	1.0141	0.1016	1.0262	0.1102	1.0659	0.1346

Table 2. Simulation results with $\rho = 0.5$.

		Full data		Proposed		Complete case	
$P(\delta = 0)$	β_0	Mean	MSE	Mean	MSE	Mean	MSE
	0.2	0.2015	0.0055	0.2047	0.0062	0.2076	0.0129
25%	0.5	0.5037	0.0134	0.5097	0.0161	0.5121	0.0334
	1.0	1.0115	0.0341	1.0183	0.0378	1.0347	0.0769
	0.2	0.2021	0.0177	0.2085	0.0211	0.2102	0.0395
75%	0.5	0.5094	0.0395	0.5112	0.0516	0.5176	0.0972
	1.0	1.0152	0.1039	1.0518	0.1407	1.0896	0.2438

From Tables 1 and 2, it is clear that the proposed estimator of β_0 is more efficient than the complete case estimator in all cases, and it makes significant improvements over the complete case estimator when the censoring is heavy (75%) and there is a high proportion of missing censoring indicators ($\rho = 0.5$). Figures 1−2 and more simulations show that the performance of the proposed estimator of Λ_0 is close to that of the "full data" estimator, and is better than that of the complete case estimator in various rates of censoring and missingness. These results suggest that our proposed estimators of β_0 and Λ_0 are more efficient than the complete case estimators, and are adequate for practical use.

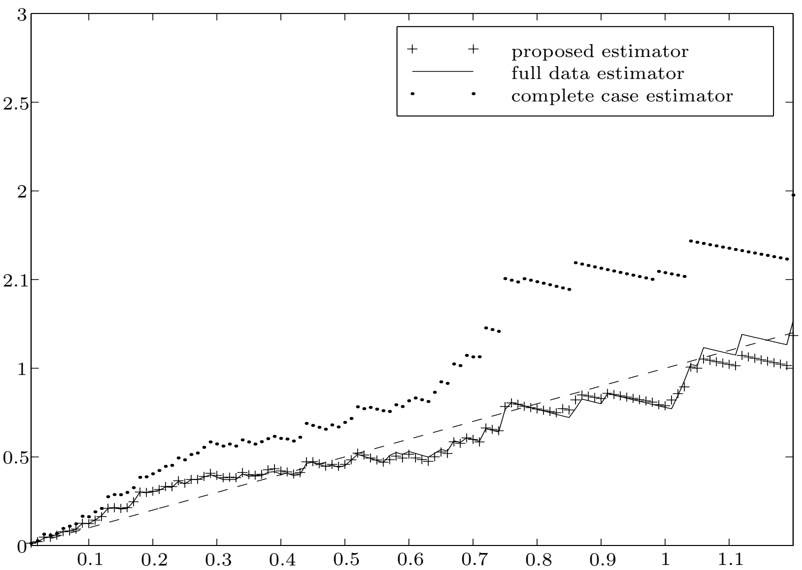

Figure 1. 25% cencoring and 20% missingness.

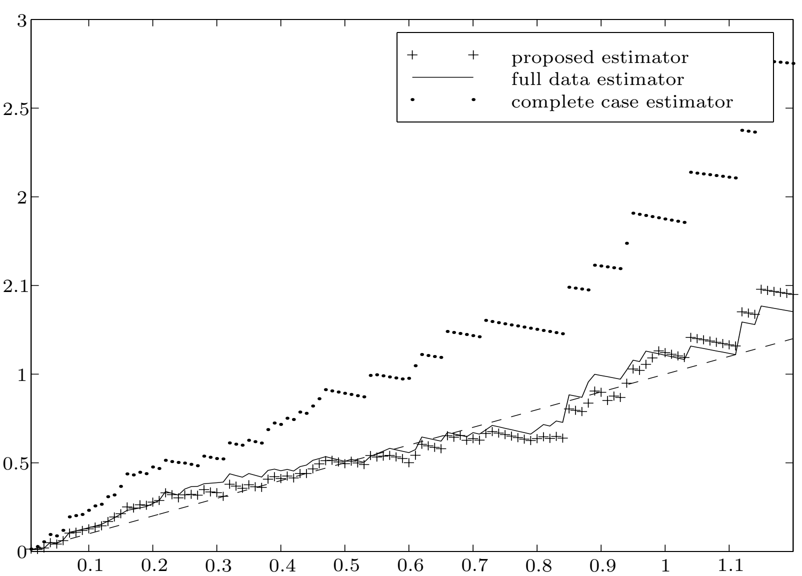

Figure 2. 25% cencoring and 50% missingness.

6. Concluding Remarks

The Cox and AH models provide two principal tools for the analyses of survival data associated with risk factors. Both models have sound biological and empirical bases, as pointed out by Breslow and Day (1987), Lin and Ying (1994), McKeague and Sasieni (1994), and so on. The choice between the two models is usually an empirical matter, since both models can provide adequate fit to any given data set if appropriate covariates are introduced. Although no method is available yet to formally test which of the two models fits the MCAR or MAR data better, some simple goodness-of-fit methods, such as plots of cumulative hazard estimates, can be used to check the adequacy of the two models.

We have proposed a simple estimation procedure for the AH model under MCAR and the case of the missingness dependent on the covariates, based on the estimating equation approach of McKeague and Subramanian (1998), and derived the asymptotic properties of the resulting estimators. As the estimator of Lin and Ying is not efficient (McKeague and Sasieni (1994)), our proposed estimators based on the estimating functions (2.6) and (4.2) may not be efficient either. However, simulation studies reported in Section 5 have shown that the estimators are reasonably good.

It may be possible to further improve the proposed estimators by other approaches, such as that of McKeague and Sasieni (1994) and the maximum likelihood methods, which point to a potential direction for further research. In addition, the estimation procedure for the general Aalen additive model (Aalen (1980)) and the semiparametric model of McKeague and Sasieni (1994) is also worth investigating.

A key assumption for the MCAR model is the assumed independence between the mechanism for missingness and everything else. This may not be appropriate in practice. We have considered the case of the missingness dependent on the covariates in Section 4. This is an important case in practice, because the availability of the censoring information (e.g., the cause of death) often depends on such covariates as age, gender, treatment method, etc. The more general MAR where the missingness may depend on other random factors, including the failure times, needs further investigations and is of considerable interest in applications.

Recently, Ibrahim, Lipsitz and Chen (1999) proposed a maximum likelihood method for estimating parameters in generalized linear models with missing covariates and a non-ignorable missing data mechanism, and used a Monte Carlo version of the EM algorithm to obtain the maximum likelihood estimates. Their likelihood approach, however, is difficult to apply to the extension considered in Section 4 because the AH model is semiparametric and the missing information is on censoring indicators instead of covariates. The estimating equation approach, on the other hand, can be easily applied to such situations. Furthermore, we should point out that the proposed missing mechanism in (4.1) is similar to the missing data mechanism considered by Ibrahim, Lipsitz and Chen (1999) and Lipsitz, Ibrahim and Zhao (1999).

7. Proofs

Proof of Theorem 1. The results are proved using the same techniques as in McKeague and Subramanian (1998), and therefore only limited details are provided here. First, a Taylor expansion of $U(\hat{\beta},\tau)$ around β_0 yields

$$n^{-1/2}U(\beta_0,\tau) = n^{-1}I(\beta^*,\tau)n^{1/2}(\hat{\beta}-\beta_0), \tag{7.1}$$

where β^* is on the line segment between $\hat{\beta}$ and β_0, and $I(\beta,\tau)$ is minus the derivative matrix of $U(\beta,\tau)$ with respect to β'. The essential part of the proof is to show the asymptotic normality of $n^{-1/2}U(\beta_0,\tau)$ and the consistency of $\hat{\beta}$. Here we only consider the case of a one-dimensional covariate, since the general case is similar. We first show the asymptotic normality of $n^{-1/2}U(\beta_0,\tau)$. Following (2.6) we can write

$$U(\beta,t) = U_{11}(\beta,t) + U_{00}(\beta,t) - \tilde{P}(\beta,t)U_{10}(\beta,t), \tag{7.2}$$

where $\tilde{P}(\beta,t) = \mathrm{diag}(U_{00}(\beta,t))[\mathrm{diag}U_{11}(\beta,t)+\mathrm{diag}U_{10}(\beta,t)]^{-1}$ and U_{00}, U_{11} and U_{10} are defined in (2.5). Note that $\sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}Y_i(s)\lambda_0(s)ds = 0$. Hence

$$\begin{aligned}
U_{11}(\beta_0,t) &= \sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}(\xi_i-\rho)[dN_i^u(s)-Y_i(s)\beta_0'Z_i(s)ds] \\
&\quad +\rho\sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}dM_i(s), \qquad (7.3)\\
U_{00}(\beta_0,t) &= \sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}(1-\xi_i)dM_i(s) \\
&\quad +\sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}(\rho-\xi_i)Y_i(s)\lambda_0(s)ds \\
&\quad +\sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}(1-\xi_i)dN_i^c(s), \qquad (7.4)
\end{aligned}$$

where $M_i(t) = N_i^u(t)-\int_0^t Y_i(s)\{d\Lambda_0(s)+\beta_0'Z_i(s)ds\}$. Write $\tilde{P}(\beta,t) = Q^{-1}(\beta,t)-1$, where

$$\begin{aligned}
Q(\beta,t) &= \sum_{i=1}^n\int_0^t\{Z_i(s)-\bar{Z}(s)\}\xi_i[dN_i(s)-Y_i(s)\beta'Z_i(s)ds] \\
&\quad \times\Big\{\sum_{i=1}^n\int_0^t(Z_i(s)-\bar{Z}(s))[dN_i(s)-Y_i(s)\beta'Z_i(s)ds]\Big\}^{-1}.
\end{aligned}$$

Using (7.3)–(7.4), we have

$$U(\beta_0, t) = A_1(t) + A_2(t), \tag{7.5}$$

where $A_1(t) = \sum_{i=1}^n \int_0^t \{Z_i(s) - \bar{Z}(s)\} dM_i(s)$ and $A_2(t) = \sum_{i=1}^n \int_0^t \{Z_i(s) - \bar{Z}(s)\}$ $(1 - \xi_i Q^{-1}(\beta_0, t)) dN_i^c(s)$.

Using the Functional Central Limit Theorem (Pollard (1990, p.53)), it can be checked that since $\sup_{0 \le t \le \tau} |Q(\beta_0, t) - \rho| = O_p(n^{-1/2})$ and

$$\sup_{0 \le t \le \tau} \left| n^{-1} \sum_{i=1}^n \xi_i \int_0^t \{Z_i(s) - \bar{Z}(s)\} dN_i^c(s) - \rho E\left[\int_0^t \{Z_i(s) - \bar{z}(s)\} dN_i^c(s) \right] \right| = o_p(1),$$

$$\begin{aligned} A_2(t) &= \sum_{i=1}^n \int_0^t \{Z_i(s) - \bar{Z}(s)\} \left(1 - \frac{\xi_i}{\rho}\right) dN_i^c(s) \\ &\quad + \sum_{i=1}^n \xi_i \int_0^t \{Z_i(s) - \bar{Z}(s)\} dN_i^c(s) \frac{Q(\beta_0, t) - \rho}{\rho Q(\beta_0, t)} \\ &= \sum_{i=1}^n \rho^{-1} (\rho - \xi_i) N_i^{CZ}(t) + n\rho^{-1} (Q(\beta_0, t) - \rho) E[N_1^{CZ}(t)] + o_p(n^{1/2}) \end{aligned}$$

uniformly on $[0, \tau]$, where $N_i^{CZ}(t)$ is defined in (2.8). Applying the delta method (Andersen, Borgan, Gill and Keiding (1993, p.109)) gives

$$\begin{aligned} &Q(\beta_0, t) - \rho \\ &= \frac{1}{E[N_1^Z(t)]} \sum_{i=1}^n (\xi_i - \rho) \int_0^t \{Z_i(s) - \bar{z}(s)\} [dN_i(s) - Y_i(s) \beta_0' Z_i(s) ds] + o_p(n^{-1/2}) \end{aligned}$$

uniformly on $[0, \tau]$, where $N_i^Z(t)$ is defined in (2.9). Therefore,

$$A_2(t) = \rho^{-1} \sum_{i=1}^n (\rho - \xi_i) \{N_i^{CZ}(t) - B(t) N_i^Z(t)\} + o_p(n^{1/2}) \tag{7.6}$$

uniformly on $[0, \tau]$.

It follows from (7.5)–(7.6) that $n^{-1/2} U(\beta_0, \tau)$ converges in distribution to a normal random vector with mean zero and variance $\Sigma + (\rho^{-1} - 1) E\{N_1^{CZ}(\tau) - B(\tau) N_1^Z(\tau)\}^{\otimes 2}$.

We next prove the consistency of $\hat{\beta}$. It follows from (7.2) that

$$-n^{-1} \frac{\partial U(\beta, \tau)}{\partial \beta'} = n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(s) - \bar{Z}(s)\} Y_i(s) Z_i'(s) ds + R_n(\beta, \tau),$$

$$
\begin{aligned}
R_n(\beta,\tau) = n^{-1}\sum_{i=1}^{n}\int_0^\tau &\{Z_i(s)-\bar{Z}(s)\}\xi_i dN_i^c(s)\\
&\times\Big\{\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}Y_i(s)Z_i'(s)ds\\
&\times\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i[dN_i(s)-Y_i(s)\beta' Z_i(s)ds]\\
&-\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i Y_i(s)Z_i'(s)ds\\
&\times\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}[dN_i(s)-Y_i(s)\beta' Z_i(s)ds]\Big\}\\
&\times\Big(\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i[dN_i(s)-Y_i(s)\beta' Z_i(s)ds]\Big)^{-2}.
\end{aligned}
$$

Using the independence assumption of the MCAR model and the Uniform Strong Law of Large Numbers (Pollard (1990, p.41)), we have, almost surely,

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i dN_i^c(s)\to \rho E N_1^{CZ}(\tau),\\
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i[dN_i(s)-Y_i(s)\beta_0' Z_i(s)ds]\to \rho E N_1^{Z}(\tau),\\
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}Y_i(s)Z_i'(s)ds\to E\Big[\int_0^\tau \{Z_1(s)-\bar{z}(s)\}Y_1(s)Z_1'(s)ds\Big],\\
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i[dN_i(s)-Y_i(s)\beta_0' Z_i(s)ds]\to \rho E N_1^{Z}(\tau),\\
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}\xi_i Y_i(s)Z_i'(s)ds\to \rho E\Big[\int_0^\tau \{Z_1(s)-\bar{z}(s)\}Y_1(s)Z_1'(s)ds\Big],\\
&n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s)-\bar{Z}(s)\}[dN_i(s)-Y_i(s)\beta_0' Z_i(s)ds]\to E N_1^{Z}(\tau).
\end{aligned}
$$

Thus, $R_n(\beta_0,\tau)=o(1)$ almost surely. Therefore,

$$
-\,n^{-1}\frac{\partial U(\beta_0,\tau)}{\partial \beta'}\to A \tag{7.7}
$$

almost surely. Note that $\partial U(\beta,\tau)/\partial\beta'$, as a function of β, is uniformly continuous in a neighborhood of β_0. Along the lines of Lin and Ying (1995, p.1717) (see also Andersen and Gill (1982) or Foutz (1977)), we conclude that $\hat{\beta}$ exists and is consistent, and that $I(\beta^*,\tau) \to A$ almost surely. This completes the proof of Theorem 1.

Proof of Theorem 2(i). Using $\Lambda_{11}^0(t) = \rho\Lambda_0(t)$, we get

$$\hat{\Lambda}_{11}^0(t) - \Lambda_{11}^0(t) = \sum_{i=1}^n \rho \int_0^t \frac{dM_i(s)}{\sum_{i=1}^n Y_i(s)} + \rho(\beta_0 - \hat{\beta})' \sum_{i=1}^n \int_0^t \frac{Y_i(s)Z_i(s)ds}{\sum_{i=1}^n Y_i(s)}$$
$$-\hat{\beta}' \sum_{i=1}^n \int_0^t \frac{(\xi_i - \rho)Y_i(s)Z_i(s)ds}{\sum_{i=1}^n Y_i(s)} + \sum_{i=1}^n \int_0^t \frac{(\xi_i - \rho)dN_i^u(s)}{\sum_{i=1}^n Y_i(s)}.$$

Note that $\sup_{0\le t\le\tau} \left| n^{-1}\sum_{i=1}^n Y_i(t) - E[Y_1(t)] \right| = O_p(n^{-1/2})$. Hence the first term in the above sum is a martingale integral with a variance function converging to 0 in probability, uniformly in t. The second term converges to 0 uniformly in t by consistency of $\hat{\beta}$. The third term tends to 0 uniformly in t as well, because its integrand converges uniformly to 0 by R3 and the Chebyshev inequality. The last term is asymptotically equivalent to

$$n^{-1} \sum_{i=1}^n \int_0^t \frac{(\xi_i - \rho)dN_i^u(s)}{E[Y_1(s)]},$$

which converges uniformly to 0 by R2 and the Law of Large Numbers. Thus $\hat{\Lambda}_{11}^0(t)$ is uniformly consistent to $\Lambda_{11}^0(t)$. Likewise, the consistency of $\hat{\Lambda}_{10}^0(t)$ and $\hat{\Lambda}_{00}^0(t)$ can be obtained. This proves part(i) of Theorem 2.

Proof of Theorem 2(ii). It follows from (3.1)$-$(3.2) that

$$\begin{aligned} n^{1/2}(\hat{\Lambda}_0(t) - \Lambda_0(t)) = {}& \left(1 + \frac{D_{00}(t)\Lambda_{10}^0(t)}{\hat{\Lambda}_{10}(t) + \hat{\Lambda}_{11}^0(t)}\right) n^{1/2}(\hat{\Lambda}_{11}^0(t) - \Lambda_{11}^0(t)) \\ & + \frac{\Lambda_{11}^0(t)}{\hat{\Lambda}_{10}(t) + \hat{\Lambda}_{11}^0(t)} n^{1/2}(\hat{\Lambda}_{00}^0(t) - \Lambda_{00}^0(t)) \\ & - \frac{D_{00}(t)\Lambda_{11}^0(t)}{\hat{\Lambda}_{10}(t) + \hat{\Lambda}_{11}^0(t)} n^{1/2}(\hat{\Lambda}_{10}^0(t) - \Lambda_{10}^0(t)), \end{aligned} \tag{7.8}$$

where $D_{kl}(t)$ $(k, l = 0, 1)$ are defined in (3.4). Along the lines of the proof of Theorem 3.1 in Lin and Ying (1995, p.1722), we have

$$\begin{aligned} n^{1/2}(\hat{\Lambda}_{11}^0(t) - \Lambda_{11}^0(t)) = {}& \rho n^{-1/2} \int_0^t \frac{\sum_{i=1}^n dM_i(s)}{E[Y_1(s)]} \\ & + \rho d(t)' A^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \{Z_i(s) - \bar{Z}(s)\} dM_i(s) \\ & + n^{-1/2} \sum_{i=1}^n (\xi_i - \rho) \int_0^t \frac{dN_i^u(s) - Y_i(s)\beta_0' Z_i(s)ds}{E[Y_1(s)]} \\ & + o_p(1) \end{aligned} \tag{7.9}$$

uniformly on $[0,\tau]$, where $d(t)$ is defined in (3.5). Applying the delta method gives

$$\begin{aligned} n^{1/2}(\hat{\Lambda}^0_{10}(t)-\Lambda^0_{10}(t)) &= \rho n^{1/2}\left(\sum_{i=1}^n\int_0^t \frac{dN_i^c(s)}{\sum_{i=1}^n Y_i(s)} - \int_0^t \frac{dE[N_1^c(s)]}{E[Y_1(s)]}\right) \\ &\quad + n^{1/2}\sum_{i=1}^n\int_0^t \frac{(\xi_i-\rho)dN_i^c(s)}{\sum_{i=1}^n Y_i(s)} \\ &= \rho n^{-1/2}\sum_{i=1}^n\int_0^t \frac{d(N_i^c(s)-E[N_1^c(s)])}{E[Y_1(s)]} \\ &\quad -\rho n^{-1/2}\sum_{i=1}^n\int_0^t \frac{(Y_i(s)-E[Y_1(s)])dE[N_1^c(s)]}{(E[Y_1(s)])^2} \\ &\quad + n^{-1/2}\sum_{i=1}^n(\xi_i-\rho)\int_0^t \frac{dN_i^c(s)}{E[Y_1(s)]} + o_p(1) \end{aligned} \tag{7.10}$$

uniformly on $[0,\tau]$. Similarly, we get

$$\begin{aligned} n^{1/2}(\hat{\Lambda}^0_{00}(t)-\Lambda^0_{00}(t)) &= (1-\rho)n^{-1/2}\sum_{i=1}^n\int_0^t \frac{dM_i^c(s)}{E[Y_1(s)]} \\ &\quad +(1-\rho)d(t)'A^{-1}n^{-1/2}\sum_{i=1}^n\int_0^\tau\{Z_i(s)-\bar{Z}(s)\}dM_i(s) \\ &\quad +n^{-1/2}\sum_{i=1}^n(\rho-\xi_i)\int_0^t \frac{dN_i(s)-Y_i(s)\beta_0'Z_i(s)ds}{E[Y_1(s)]} \\ &\quad +(1-\rho)n^{-1/2}\sum_{i=1}^n\int_0^t \frac{d(N_i^c(s)-E[N_1^c(s)])}{E[Y_1(s)]} \\ &\quad -(1-\rho)n^{-1/2}\sum_{i=1}^n\int_0^t \frac{(Y_i(s)-E[Y_1(s)])dE[N_1^c(s)]}{(E[Y_1(s)])^2} \\ &\quad +o_p(1) \end{aligned} \tag{7.11}$$

uniformly on $[0,\tau]$. In view of (7.8)–(7.11) and the proof of Theorem 2(i), we obtain

$$n^{1/2}(\hat{\Lambda}_0(t)-\Lambda_0(t)) = \sum_{i=1}^4 S_{ni}(t) + o_p(1) \tag{7.12}$$

uniformly on $[0,\tau]$, where

$$\begin{aligned} S_{n1}(t) &= \Big(1+\rho D_{00}(t)D_{10}(t)+(1-\rho)D_{11}(t)\Big)n^{-1/2}\sum_{i=1}^n\Big[\int_0^t \frac{dM_i(s)}{E[Y_1(s)]} \\ &\quad +d(t)'A^{-1}\int_0^\tau\{Z_i(s)-\bar{Z}(s)\}dM_i(s)\Big], \end{aligned}$$

$$
\begin{aligned}
S_{n2}(t) &= \Big(1 + D_{00}(t)D_{10}(t) - D_{11}(t)\Big) \\
&\qquad \times n^{-1/2}\sum_{i=1}^{n}(\xi_i - \rho)\int_0^t \frac{dN_i^u(s) - Y_i(s)\beta_0' Z_i(s)ds}{E[Y_1(s)]}, \\
S_{n3}(t) &= -D_{11}(t)\{1 + D_{00}(t)\}n^{-1/2}\sum_{i=1}^{n}(\xi_i - \rho)\int_0^t \frac{dN_i^c(s)}{E[Y_1(s)]}, \\
S_{n4}(t) &= D_{11}(t)\Big(1 - \rho - \rho D_{00}(t)\Big)n^{-1/2}\sum_{i=1}^{n}\Big\{\int_0^t \frac{d(N_i^c(s) - E[N_1^c(s)])}{E[Y_1(s)]} \\
&\qquad - \int_0^t \frac{(Y_i(s) - E[Y_1(s)])dE[N_i^c(s)]}{(E[Y_1(s)])^2}\Big\}.
\end{aligned}
$$

Since $\sum_{i=1}^4 S_{ni}(t)$ is a sum of i.i.d. processes, the convergence of the finite dimensional distribution of $\sum_{i=1}^4 S_{ni}(t)$ follows from the Multivariate Central Limit Theorem. To prove that $\sum_{i=1}^4 S_{ni}(t)$ is tight, it suffices to show the tightness for each $S_{ni}(t)$, $i = 1, 2, 3, 4$. First, $S_{n1}(t)$ is tight as it is a martingale integral. Next, note that $Z_i(t) = \max\{Z_i(t), 0\} - \max\{-Z_i(t), 0\}$, so we can also write $\int_0^t Y_i(s)\beta_0' Z_i(s)ds/E[Y_1(s)]$ as the sum of two monotone processes on $[0, \tau]$. It follows that $(\xi_i - \rho)\int_0^t\{dN_i^u(s) - Y_i(s)\beta_0' Z_i(s)ds\}/E[Y_1(s)]$ can be written as sums of monotone processes, and the tightness of $S_{n2}(t)$ follows from Example 2.11.16 of van der Vaart and Wellner (1996). Furthermore, for each i, the process $(\xi_i - \rho)\int_0^t dN_i^c(s)/E[Y_1(s)]$ has mean zero and can be expressed as the sum of two monotone processes on $[0, \tau]$. Thus, $S_{n3}(t)$ is also tight. Finally, the tightness of $S_{n4}(t)$ follows from some basic properties of empirical processes (Shorack and Wellner (1986, p.109)). This proves the weak convergence of Theorem 2(ii). Moreover, a straightforward calculation yields (3.3), which completes the proof of Theorem 2(ii).

Proof of Theorem 3. As in the proof of Theorem 1, we can write

$$U^*(\beta_0, t) = A_1(t) + A_2^*(t) + A_3^*(t), \tag{7.13}$$

where $A_1(t)$ is defined in (7.5), and

$$
\begin{aligned}
A_2^*(t) &= \sum_{i=1}^{n}\int_0^t\{Z_i - \bar{Z}(s)\}(1 - \frac{\xi_i}{\rho_i Q^*(\beta_0, t)})dN_i^c(s), \\
A_3^*(t) &= \sum_{i=1}^{n}\int_0^t\{Z_i - \bar{Z}(s)\}\frac{\xi_i}{Q^*(\beta_0, t)}(\frac{1}{\rho_i} - \frac{1}{\hat{\rho}_i})dN_i^c(s), \\
Q^*(\beta, t) &= \sum_{i=1}^{n}\int_0^t\{Z_i - \bar{Z}(s)\}\frac{\xi_i}{\hat{\rho}_i}[dN_i(s) - Y_i(s)\beta' Z_i ds] \\
&\qquad \times\Big\{\sum_{i=1}^{n}\int_0^t(Z_i - \bar{Z}(s))[dN_i(s) - Y_i(s)\beta' Z_i ds]\Big\}^{-1}.
\end{aligned}
$$

Note that $\hat{\alpha} - \alpha_0 = O_p(n^{-1/2})$ and ξ_i is conditionally independent of (T_i, C_i) given Z_i. Using the Taylor expansion of $\phi(Z_i, \hat{\alpha})$ around α_0, we obtain

$$\sup_{0 \le t \le \tau} |Q^*(\beta_0, t) - 1| = O_p(n^{-1/2}) \tag{7.14}$$

and

$$\sup_{0 \le t \le \tau} \left| n^{-1} \sum_{i=1}^n \frac{\xi_i}{\rho_i} \int_0^t \{Z_i - \bar{Z}(s)\} dN_i^c(s) - E\left[\int_0^t \{Z_i - \bar{z}(s)\} dN_i^c(s) \right] \right| = o_p(1).$$

Thus,

$$\begin{aligned} A_2^*(t) &= \sum_{i=1}^n \int_0^t \{Z_i - \bar{Z}(s)\} \left(1 - \frac{\xi_i}{\rho_i}\right) dN_i^c(s) \\ &\quad + \sum_{i=1}^n \frac{\xi_i}{\rho_i} \int_0^t \{Z_i - \bar{Z}(s)\} dN_i^c(s) \frac{Q^*(\beta_0, t) - 1}{Q^*(\beta_0, t)} \\ &= \sum_{i=1}^n \rho_i^{-1} (\rho_i - \xi_i) N_i^{CZ}(t) + n\rho_i^{-1} (Q^*(\beta_0, t) - 1) E[N_1^{CZ}(t)] + o_p(n^{1/2}) \end{aligned}$$

uniformly on $[0, \tau]$. Note that $Q^*(\beta_0, t)$ in the denominator above can be replaced with 1 in view of (7.14). In addition, the delta method gives

$$Q^*(\beta_0, t) - 1 = \frac{1}{E[N_1^Z(t)]} \sum_{i=1}^n \rho_i^{-1} (\xi_i - \rho_i) N_i^Z(t) + o_p(n^{-1/2})$$

uniformly on $[0, \tau]$. Therefore,

$$A_2^*(t) = \sum_{i=1}^n \rho_i^{-1} (\rho_i - \xi_i) \{N_i^{CZ}(t) - B(t) N_i^Z(t)\} + o_p(n^{1/2}) \tag{7.15}$$

uniformly on $[0, \tau]$. Note that each ξ_i is a binary variable. Thus it follows from likelihood theory that

$$\hat{\alpha} - \alpha_0 = \Gamma^{-1} \frac{1}{n} \sum_{i=1}^n \frac{\xi_i - \rho_i}{\rho_i (1 - \rho_i)} \frac{\partial \phi(Z_i, \alpha_0)}{\partial \alpha} + o_p(n^{-1/2}). \tag{7.16}$$

Using (7.14), (7.16) and the delta method, we have

$$A_3^*(t) = E\left(\rho_1 N_1^{CZ}(t) \frac{\partial \phi(Z_1, \alpha_0)}{\partial \alpha'}\right) \Gamma^{-1} \sum_{i=1}^n \frac{\xi_i - \rho_i}{\rho_i (1 - \rho_i)} \frac{\partial \phi(Z_i, \alpha_0)}{\partial \alpha} + o_p(n^{1/2}). \tag{7.17}$$

It follows from (7.13), (7.15) and (7.17) that $n^{-1/2} U^*(\beta_0, \tau)$ converges in distribution to a zero mean normal random vector with variance $\Sigma + \Phi$. Similar to

(7.7), we get $-n^{-1}\partial U^*(\beta_0,\tau)/\partial\beta' \to A$ almost surely. Consequently, the results of Theorem 3 follow from the same arguments as in the proof of Theorem 1.

Acknowledgement

We are grateful to an associate editor and the referee for their valuable comments and suggestions that have contributed greatly to the improvement of the paper. This work is partially supported by Internal Research Grant No.A-PC37 of The Hong Kong Polytechnic University and Grant No. 10071092 of the National Science Foundation of China.

References

Aalen, O. O. (1982). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory* (Edited by Klonecki W., Kozek A. and Rosinski J.), 1-25. Lecture Notes in Statistics **2**, Springer-Verlag, New York.

Andersen, P. K., Borgan, Φ., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Andersen, P. K. and Gill. R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research*, **2**, The Design and Analysis of Cohort studies. IARC, Lyon.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* **38**, 417-431.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Amer. Statist. Assoc.* **72**, 147-148.

Gijbels, I., Lin, D. and Ying, Z. (1993). Non- and semi-parametric analysis of failure time data with missing failure indicators. Tech. Report 039-93, Mathematical Sciences Research Institute, Berkeley.

Gill, R. D., van der Laan, M. J. and Robins, J. M. (1997). Coarsening at random. In 1*st Seattle Symp. Biostatist.* (Edited by D.-Y. Lin). Springer, New York.

Goetghebeur, E. J. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* **82,** 821-833.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244-2253.

Ibrahim, J. G., Lipsitz, S. R. and Chen, M. H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J. Roy. Statist. Soc. Ser. B* **61,** 173-190.

Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *Ann. Statist.* **23**, 774-786.

van der Laan, M. J. and McKeague, I. W. (1998). Efficient estimation from right-censored data when failure indicators are missing at random. *Ann. Statist.* **26**, 164-182.

Lin, Y. D. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

Lin, Y. D. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard risk models for counting processes. *Ann. Statist.* **23**, 1712-1734.

Lipsitz, S. P., Ibrahim, J. G. and Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *J. Amer. Statist. Assoc.* **94,** 1147-1160.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. Wiley, New York.

Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data. *J. Multivariate Anal.* **39**, 217-235.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501-514.

McKeague, I. W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information. *Scand. J. Statist.* **25**, 589-601.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, Hayward, California.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-590.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Subramanian, S. (2000). Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazards. *J. Statist. Plann. Inference* **84**, 81-94.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Spring-Verlag, New York.

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

E-mail: maxzhou@polyu.edu.hk

Institute of Applied Mathematics, Academia Sinica, Beijing, P.R. China

(Received January 2001; accepted February 2003)