Statistica Sinica **20** (2010), 1125-1142

SEMIPARAMETRIC ESTIMATION FOR REGRESSION COEFFICIENTS IN THE COX MODEL WITH FAILURE INDICATORS MISSING AT RANDOM

Chunling Liu[3] and Qihua Wang[1,2,3]

[1] *Yunnan University,* [2] *Chinese Academy of Sciences and* [3] *The University of Hong Kong*

Abstract: This paper considers a regression imputation method for estimating the regression coefficients in the Cox model when some failure indicators are missing at random, and the conditional probability of the censoring indicator is assumed to be of a parametric form. To avoid problems with missspecification of the parametric form, two augmented inverse probability weighted estimators are defined, and their asymptotic properties are established. Simulation studies were conducted to demonstrate the performance of the proposed estimators, and a data set from a stage II breast cancer trial is used to illustrate our methods.

Key words and phrases: Augmented inverse probability weighting, Cox proportional hazards model, missing at random, Nadaraya-Watson kernel estimate, regression imputation.

1. Introduction

Failure time T, associated with a p-vector Z of possibly time-varying covariates, is commonly characterized by the proportional hazards regression model as follows:

$$\lambda(t|Z) = \lambda_0(t)\exp\{\beta_0^\top Z(t)\},$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β is a p-vector of unknown regression coefficients, see Cox (1972). Let C denote the censoring time and $I(\cdot)$ be the indicator function. Under random censorship, one observes only (X,δ,Z), where $X = \min\{T,C\}$ and $\delta = I(T \le C)$. Assume that the censoring is noninformative in the sense that T and C are independent given Z. Andersen and Gill (1982) made elegant use of counting processes and martingale theory to obtain the asymptotic properties of the maximum partial likelihood estimate for the regression coefficients β, and hence the baseline hazard function $\lambda_0(t)$.

The failure indicator δ may not always be available. For instance, for competing risks survival data, failure is attributed to multiple causes. If one just assesses the effects of covariates on cause-specific hazard, failure times beyond interest are also treated as censoring variables. Meanwhile when cause of failure is unknown, as might happen in some settings, to save expense without autopsy say, δ is missing. For example, van der Laan and Mckeague (1998) describe epidemiological studies in which death certificates were missing for some people mainly due to emigration, or inconclusive hospital case notes and autopsy results, and point out that it may be impossible to determine whether death was due to the cause of interest. Missing causes of death can also arise in some studies when only a subset of animals is examined for tumors to cut costs. Occasionally, tissues autolyze or are cannibalized by cage mates before a necropsy can be performed, and pathologists are not always able to determine each tumor's role in causing death. Kalbfleisch and Prentice (1980) provide data on mice who died from leukemia, from other known (non-leukemia) causes, or from unknown causes; and Cummings et al. (1986) present a data set on elderly women with breast cancer who died from breast cancer, from other known causes, or from unknown causes. This last data set is analyzed in Section 5.

A naive method for handling missing data is the complete case (CC) method, i.e., discard records with unknown failure indicators. Such a strategy is easy to carry out but tends to be highly inefficient for small sample cases, and to be biased for dependently missing mechanisms, see Little and Rubin (1987). Some authors improved on the CC method at the expense of invoking the missing completely at random (MCAR) assumption. Thus, Gijbels, Lin and Ying (1993) constructed an estimating equation for subjects with missing failure indictors based on the estimated missing proportion, and hence derived an efficient estimator for the vector of regression coefficients. McKeague and Subramanian (1998) considered an alternative estimating equation based on certain cumulative transition intensities, and discussed the asymptotic properties of the conditional survival function; the asymptotic efficiencies of the estimates were discussed by Subramanian (2000) for the case of MCAR and the case without missing. Zhou and Sun (2003) extended McKeague and Subramanian (1998) to the additive hazard regression model, but MCAR was also assumed. The method of Gijbels, Lin and Ying (1993) depends heavily on the assumption of MCAR, and McKeague and Subramanian (1998) also mentioned that the missing mechanism in their work could not be relaxed without modification. However, the MCAR assumption is impractical in many cases.

Missing at random (MAR) is a more general assumption that is commonly used in missing data analysis. Here missingness depends on observed data only. There has been a good deal of work on the framework of competing risks model with the failure types missing at random, see Goetghebeur and Ryan (1995), Lu and Tsiatis (2001), Gao and Tsiatis (2005), among others. The existing estimators due to Gijbels, Lin and Ying (1993) and McKeague and Subramanian (1998) for the Cox model are inconsistent under MAR.

In this paper, we treat the estimation of Cox models with failure indicators missing at random. In Section 2, we specify a parametric form for the conditional mean function of censoring given covariates, and derive an estimator for the vector of regression coefficients by performing a regression imputation method. To avoid problems with the misspecified parametric forms, in Section 3, we further consider an augmented inverse probability weighted (AIPW) estimation for the parameter vector β, see Robins, Rotnitzky and Zhao (1994), Wang et al. (1997) and Tsiatis (2006), while the probability of missingness is fitted by a nonparametric method and a parametric one. Some simulation experiments are presented in Section 4, and we illustrate the results via a data set from a clinical trial involving elderly women with stage II breast cancer in Section 5. The proofs of main results are given in Appendices A and B.

2. Regression Imputation Estimation

Let $\{X_i, \delta_i, Z_i(t)\}$, $i = 1, \ldots, n$, be n independent copies of $\{X, \delta, Z(t)\}$. When failure indicators are observed completely, the maximum partial likelihood estimate of β_0 is obtained as the solution of the following estimation equation

$$U(\beta) := \sum \delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Y_j(X_i) \exp(\beta^\top Z_j(X_i)) Z_j(X_i)}{\sum_{j=1}^{n} Y_j(X_i) \exp(\beta^\top Z_j(X_i))} \right\} = 0, \tag{2.1}$$

where $Y_i(\cdot) = I(X_i \geq \cdot)$ is the at-risk process, see Cox (1975). Andersen and Gill (1982) considered this estimator, denoted by $\hat{\beta}_{AG}$, and established asymptotic normality by using counting processes and martingale theory. For notational convenience, denote $\{X_i, Z_i(X_i)\}$ by W_i, and the ratio under the brackets in (2.1) by $\bar{Z}(\beta, X_i)$.

Let ξ be an indicator, it is set to zero when δ is missing and is one otherwise. Hence the observed data are $\{W_i, \delta_i, \xi_i = 1\}$ or $\{W_i, \xi_i = 0\}$. For a missing failure indicator, say δ_i, its conditional expectation $\mathbb{E}(\delta_i | W_i) = m(W_i)$ is a natural surrogate, and then (2.1) can be used to obtain the estimate of β by replacing the missing δ_i by an estimate of $m(W_i)$. We posit a parametric model $m_0(W_i, \theta_0)$ for $m(W_i)$ with unknown parameter vector θ_0 that can be estimated by a maximum likelihood approach. The conditional mean function $m(\cdot)$ may then be estimated parametrically through $m_0(\cdot, \hat{\theta}_n)$, where $\hat{\theta}_n$ is the maximizer of the likelihood function

$$L_n(\theta) = \prod m_0(W_i, \theta)^{\xi_i \delta_i} [1 - m_0(W_i, \theta)]^{\xi_i (1-\delta_i)}.$$

The widely used logit model may be used for $m_0(W_i, \theta)$ since δ_i is binary, and other parametric forms are possible. An imputation estimator, denoted by $\widetilde{\beta}_n$, can then be defined to be the solution in β of

$$\widetilde{U}(\beta) = \sum [\xi_i \delta_i + (1-\xi_i) m_0(W_i, \hat{\theta}_n)] \Big\{ Z_i - \bar{Z}(\beta, X_i) \Big\} = 0. \tag{2.2}$$

For stating the asymptotic behavior of $\widetilde{\beta}_n$, we need some notation as follows:

$$\begin{aligned}
\nabla m_0(W_i, \theta) &= \frac{\partial m_0(W_i, \theta)}{\partial \theta}; \\
I(\theta) &= \mathbb{E} \frac{\xi_i \nabla m_0(W_i, \theta) \nabla^\top m_0(W_i, \theta)}{m_0(W_i, \theta)\{1 - m_0(W_i, \theta)\}}; \\
S^{(r)}(\beta, t) &= n^{-1} \sum Y_i(t) \exp(\beta^\top Z_i(t)) Z_i(t)^{\bigotimes r}, \ r = 0, 1, 2; \\
s^{(r)}(\beta, t) &= \mathbb{E}\{S^{(r)}(\beta, t)\}, \ r = 0, 1, 2; \\
\bar{Z}(\beta, t) &= \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}, \quad \bar{z}(\beta, t) = \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)}; \\
\Sigma_1 &= \mathbb{E}\{[Z_1 - \bar{z}(\beta_0, X_1)]^{\bigotimes 2} m_0(W_1, \theta_0)\}; \\
\Sigma_2 &= \mathbb{E}\{[Z_1 - \bar{z}(\beta_0, X_1)]^{\bigotimes 2} m_0(W_1, \theta_0)[1 - m_0(W_1, \theta_0)](1 - \xi_1)\}; \\
\Sigma_3 &= \mathbb{E}\{[Z_1 - \bar{z}(\beta_0, X_1)] \nabla^\top m_0(W_1, \theta_0)\}; \\
\Sigma_4 &= \mathbb{E}\{[Z_1 - \bar{z}(\beta_0, X_1)] \nabla^\top m_0(W_1, \theta_0) \xi_1\}; \\
\Sigma_0 &= \Sigma_1 - \Sigma_2 + \Sigma_3 I^{-1}(\theta_0) \Sigma_3^\top - \Sigma_4 I^{-1}(\theta_0) \Sigma_4^\top.
\end{aligned}$$

Note that $\bar{Z}(\beta, X_i)$ is just the ratio in the brackets at (2.1). Let $\tau_H = \sup\{t : 1 - H(t) > 0\}$ and τ_0 be a positive constant, where $H(t)$ is the survival function of X. Two assumptions are needed:

(C1) $|Z(t)| \leq K < \infty$ a.s.;

(C2) the function $H(t)$ is continuous on $[\tau_0, \tau_H)$, and the matrix $\Sigma_0(\tau)$ given in Appendix A is positive definite for each $\tau \in [\tau_0, \tau_H)$.

The asymptotic normality of $\widetilde{\beta}_n$ can now be stated.

Theorem 2.1.

(i) *Under* (C1) *and* (C2), *if* (A1)$-$(A6) *in Appendix A are satisfied, then* $n^{-1/2} \widetilde{U}(\beta_0) \to_d \mathcal{N}(0, \Sigma_0)$.

(ii) *If* Σ_1 *is positive definite,* $n^{1/2}(\widetilde{\beta}_n - \beta_0) \to_d \mathcal{N}(0, \Sigma_1^{-1} \Sigma_0 \Sigma_1^{-1})$.

When all failure indicators are fully observed, i.e., all ξ_i's are equal to one, then $\Sigma_2 = 0$, $\Sigma_3 = \Sigma_4$, so $\Sigma_0 = \Sigma_1$ and one has the results in Andersen and Gill

(1982). Let

$$\begin{aligned}
\hat{\Sigma}_1 &= n^{-1}\sum[Z_i - \bar{Z}(\tilde{\beta}_n, X_i)]^{\bigotimes 2} m_0(W_i, \hat{\theta}_n),\\
\hat{\Sigma}_2 &= n^{-1}\sum[Z_i - \bar{Z}(\widetilde{\beta}_n, X_i)]^{\bigotimes 2} m_0(W_i, \hat{\theta}_n)[1 - m_0(W_i, \hat{\theta}_n)][1 - \xi_i],\\
\hat{\Sigma}_3 &= n^{-1}\sum[Z_i - \bar{Z}(\widetilde{\beta}_n, X_i)]\nabla^\top m_0(W_i, \hat{\theta}_n),\\
\hat{\Sigma}_4 &= n^{-1}\sum[Z_i - \bar{Z}(\widetilde{\beta}_n, X_i)]\nabla^\top m_0(W_i, \hat{\theta}_n)\xi_i\\
\hat{\Sigma}_0 &= \hat{\Sigma}_1 - \hat{\Sigma}_2 + \hat{\Sigma}_3 I^{-1}(\hat{\theta}_n)\hat{\Sigma}_3^\top - \hat{\Sigma}_4 I^{-1}(\hat{\theta}_n)\hat{\Sigma}_4^\top.
\end{aligned}$$

It can be shown that $\hat{\Sigma}_i$, $i = 1, \ldots, 4$ are consistent estimators for Σ_i, $i = 1, \ldots, 4$, respectively. Thus $\hat{\Sigma}_0$ is also a consistent estimator for Σ_0.

3. Augmented Inverse Probability Weighted Estimation

The regression imputation estimator in Section 2 may be biased when the parametric form $m_0(W_i, \theta)$ is misspecified. The augmented inverse probability weighted (AIPW) method is commonly used to solve this problem, and such estimates enjoy the property of double robustness, see Wang et al. (1997), Robins, Rotnitzky and Zhao (1994), and Tsiatis (2006).

Under the MAR assumption, the indicators ξ_i and δ_i are independent conditional on W_i, i.e., $\mathbb{P}(\xi_i = 1|W_i, \delta_i) = \mathbb{P}(\xi_i = 1|W_i) = \pi(W_i)$. Generally, the function $\pi(\cdot)$ is estimated by assuming a parametric model for $\pi(\cdot)$, and then acts as the weights in AIPW methods. However, the resulting estimator is inconsistent when both $\pi(\cdot)$ and $m(\cdot)$ are misspecified. This motivates us to use a nonparametric method to estimate $\pi(\cdot)$ to obtain a robust estimation of the regression coefficients. This can be done by using the Nadaraya-Watson kernel in the form

$$\pi_n(w) = \frac{\sum \xi_i \mathcal{K}_H(w - W_i)}{\sum \mathcal{K}_H(w - W_i)},$$

where $\mathcal{K}_H(\cdot) = \{\det(H)\}^{-1}\mathcal{K}(H^{-1}\cdot)$ denotes a multivariate kernel function with bandwidth matrix H. The bandwidth matrix can be chosen by Scott's rule, see Scott (1992). Under some regularity conditions, $\pi_n(w)$ is consistent, i.e.,

$$\sup_w |\pi_n(w) - \pi(w)| = o_p(1), \tag{3.1}$$

see Devroye (1978) and Hardle, Janssen and Serfling (1988). Then we derive a new estimating equation as,

$$S(\beta) = \sum\left\{\frac{\xi_i\delta_i}{\pi_n(W_i)} + \Big[1 - \frac{\xi_i}{\pi_n(W_i)}\Big]m_0(W_i, \hat{\theta}_n)\right\}\{Z_i - \bar{Z}(\beta, X_i)\} = 0, \tag{3.2}$$

and the augmented inverse probability weighted estimator $\tilde{\beta}_n^{IPW}$ is defined as the solution to (3.2).

We now show the consistency of $\tilde{\beta}_n^{IPW}$. It is sufficient to prove that the expectation of (3.2) with $[\pi_n(W_i), \beta, \hat{\theta}_n]$ replaced by $[\pi(W_i), \beta_0, \theta^*]$ is zero, see Gao and Tsiatis (2005). Note then that

$$\sum[Z_i - \bar{Z}(\beta_0, X_i)]\left[\frac{\xi_i\delta_i}{\pi(W_i)} + \left\{1 - \frac{\xi_i}{\pi(W_i)}\right\}m_0(W_i, \theta^*)\right]$$
$$= \sum[Z_i - \bar{Z}(\beta_0, X_i)]\delta_i + \sum[Z_i - \bar{Z}(\beta_0, X_i)]\left[1 - \frac{\xi_i}{\pi(W_i)}\right][m_0(W_i, \theta^*) - \delta_i]. \quad (3.3)$$

The expectation of the first term on the right side of (3.3) is readily shown to be zero by martingale theory for counting processes (Fleming and Harrington (1991)), and the expectation value of the second summation in (3.3) is always zero regardless of $\theta^* = \theta_0$. Thus, the augmented inverse probability weighted estimator $\tilde{\beta}_n^{IPW}$ is consistent, hence robust.

To construct the inverse probability weighted estimate, a technical condition is needed,

(C3) $\inf_w \pi(w) > 0$.

For any $\tau < \tau_H$, let

$$\Sigma_5(\tau) = \mathbb{E}\Big\{[Z_1 - \bar{z}(\beta_0, X_1)]^{\bigotimes 2} m_0(W_1, \theta_0)[1 - m_0(W_1, \theta_0)][\pi^{-1}(W_1) - 1]I(X_1 \le \tau)\Big\}.$$

Let $\Sigma_5 = \Sigma_5(\tau_H)$. In order to state the asymptotic normality of $\tilde{\beta}_n^{IPW}$, we need another assumption.

(C4) There exists a $\tau_0 < \tau_H$ such that H is continuous on $[\tau_0, \tau_H)$, and the matrix $\Sigma_1(\tau)$, given in Appendix A, or $\Sigma_5(\tau)$ is positive definite for each $\tau \in [\tau_0, \tau_H)$.

Theorem 3.1.

(i) *Under* (C1), (C3) *and* (C4), *if* (A1)−(A6) *in Appendix A and* (3.1) *are satisfied, then* $n^{-1/2}S(\beta_0) \to_d \mathcal{N}(0, \Sigma_1 + \Sigma_5)$.

(ii) *If* Σ_1 *is positive definite, then* $n^{1/2}(\tilde{\beta}^{IPW} - \beta_0) \to_d \mathcal{N}(0, \Sigma_1^{-1} + \Sigma_1^{-1}\Sigma_5\Sigma_1^{-1})$.

For survival data, if we know the missing mechanism well, a parametric method might be used to fit the probability of missingness. Suppose $\pi(w) = \pi_0(w, \gamma_0)$, where the form of the function $\pi_0(w, \gamma)$ is known, and depends on an unknown finite dimensional parameter vector γ. We estimate γ_0 by $\hat{\gamma}_n$, the maximizer of the likelihood function,

$$\prod \pi_0(W_i, \gamma)^{\xi_i}[1 - \pi_0(W_i, \gamma)]^{1-\xi_i}.$$

Under some regularity conditions, the estimator $\hat{\gamma}_n$ satisfies

$$\hat{\gamma}_n - \gamma_0 = O_p(n^{-1/2}). \tag{3.4}$$

Based on the above estimation, we replace $S(\beta)$ by

$$S^*(\beta) = \sum \left[\frac{\xi_i \delta_i}{\pi_0(W_i, \hat{\gamma}_n)} + \left\{ 1 - \frac{\xi_i}{\pi_0(W_i, \hat{\gamma}_n)} \right\} m_0(W_i, \hat{\theta}_n) \right] [Z_i - \bar{Z}(\beta, X_i)] = 0,$$

and we define the new augmented inverse probability weighted estimator $\tilde{\beta}_n^{DR}$ to be its solution in β. Note that $\tilde{\beta}_n^{DR}$ is double robust, see Gao and Tsiatis (2005), and asymptotically normal.

Theorem 3.2. *Under* (C1), (C3), (C4), (A1)−(A6), *and* (3.4), *if* Σ_1 *is positive definite, then* $n^{1/2}(\tilde{\beta}^{DR} - \beta_0) \to_d \mathcal{N}(0, \Sigma_1^{-1} + \Sigma_1^{-1}\Sigma_5\Sigma_1^{-1})$.

Note that $\widetilde{\beta}^{IPW}$ and $\widetilde{\beta}^{DR}$ have the same asymptotic variance. We recommend use of $\tilde{\beta}^{IPW}$ if the dimension of w is low, and $\tilde{\beta}^{DR}$ otherwise.

We conclude that the methods of fitting $\pi(W_i)$ asymptotically have no effect on the inverse probability weighted estimates. Then if

$$\hat{\Sigma}_5 = n^{-1} \sum [Z_i - \bar{Z}(\widetilde{\beta}_n^*, X_i)]^{\bigotimes 2} m_0(W_i, \hat{\theta}_n)[1 - m_0(W_i, \hat{\theta}_n)][\widetilde{\pi}^{-1}(W_i) - 1],$$

where $(\widetilde{\beta}_n^*, \widetilde{\pi}(W_i))$ can be either of $(\widetilde{\beta}^{IPW}, \pi_n(W_i))$ or $(\widetilde{\beta}^{DR}, \pi_0(W_i, \hat{\gamma}))$, $\hat{\Sigma}_5$ is a consistent estimator of Σ_5, and the matrix $\hat{\Sigma}_1^{-1} + \hat{\Sigma}_1^{-1}\hat{\Sigma}_5\hat{\Sigma}_1^{-1}$ can be used to estimate the asymptotic variance.

4. Simulation Studies

Monte Carlo simulations were conducted to evaluate the finite-sample behavior of the proposed estimators $\widetilde{\beta}_n$, $\widetilde{\beta}_n^{IPW}$, and $\widetilde{\beta}^{DR}$, and to compare them with the full-data estimator $\hat{\beta}_{AG}$ and the complete-case estimator $\hat{\beta}_{CC}$. Although it is not achievable because of the missingness of failure indicators, $\hat{\beta}_{AG}$ can serve as a gold standard.

The baseline hazard $\lambda_0(t)$ was set at one, and a univariate covariate Z was uniformly distributed on [0, 2]. The censoring variable was generated independent of Z_i from an exponential distribution with rate λ_c. Under this scenario, the true $m(w)$ is a logistic regression model $1/[1 + \lambda_c \exp(-\beta_0 z)]$. We set $m_0(w, \theta) = 1/[1 + \exp(-\theta_1 - \theta_2 x - \theta_3 z)]$, and λ_c varied to yield censoring rates of approximately 30% and 70%. A logistic regression $\pi_0(w, a) = 1/[1+\exp(-a_0-a_1x-a_2z)]$ was adopted and $(a_0, a_1, a_2)^\top$ was adjusted to generate missing rates of roughly 20% and 50%. A bivariate Epanechnikov kernel function was chosen for the bivariate Nadaraya-Watson estimate of the probability of missingness. As to the

Table 1. Simulation results comparing five estimators for different combinations of censoring rate (CR) and missing rate (MR) (1, 000 replications and $n = 200$).

MR	CR		$\hat{\beta}_{AG}$	$\hat{\beta}_{CC}$	$\widetilde{\beta}_n$	$\widetilde{\beta}_n^{IPW}$	$\widetilde{\beta}^{DR}$
20%	30%	Bias	-0.0005	0.1872	0.0012	0.0009	0.0008
		SSE	0.1558	0.1764	0.1636	0.1654	0.1647
		SEE	0.1534	0.1752	0.1614	0.1799	0.1623
		CP	0.947	0.825	0.946	0.957	0.944
	70%	Bias	0.0041	-0.2553	0.0034	0.0046	0.0058
		SSE	0.1525	0.2142	0.1726	0.1762	0.1752
		SEE	0.1534	0.2227	0.1757	0.1776	0.1780
		CP	0.959	0.792	0.957	0.959	0.948
50%	30%	Bias	0.0097	0.2409	0.0148	0.0176	0.0174
		SSE	0.2411	0.2679	0.2677	0.2743	0.2723
		SEE	0.2361	0.2722	0.2614	0.2622	0.2649
		CP	0.948	0.871	0.950	0.959	0.949
	70%	Bias	0.0087	-0.1356	0.0026	0.0013	0.0008
		SSE	0.2359	0.3389	0.3064	0.3130	0.3120
		SEE	0.2359	0.3439	0.3097	0.3168	0.3126
		CP	0.957	0.936	0.959	0.958	0.959

SSE: Estimated standard error; SEE: Monte Carlo empirical standard error; CP: Empirical coverage probability.

parametric estimate of the probability of missingness, a logit model was adopted again.

We chose $\beta_0 = 0.5$. The sample size was set to be 200. For each scenario, the Monte Carlo simulation consisted of 1,000 replicates. Tables 1 presents the bias, the estimated standard errors (SSE), the Monte-Carlo empirical standard errors (SEE), and the empirical coverage probability (CP) of the 95% confidence interval.

Table 1 summarizes the results for various scenarios of censoring rates and missing rates. The simulation results were consistent with the theoretical ones. In all cases, three presented estimates are significantly superior to the complete-case estimator as to accuracy or efficiency. The bias of the augmented inverse probability estimates were always close to zero (consistency). The SEE was very close to the SSE, except for the case of estimating the probability of missingness nonparametrically. The bias, SSE, and SEE of the proposed estimators had good efficiency compared to $\hat{\beta}_{AG}$. Furthermore, in almost all cases, the coverage probabilities of the normal approximation confidence intervals of the three proposed estimators were close to 95%.

Table 2. Simulation results with the parametric form of m misspecified.

MR	CR		$\hat{\beta}_{AG}$	$\hat{\beta}_{CC}$	$\widetilde{\beta}_n$	$\widetilde{\beta}_n^{IPW}$	$\widetilde{\beta}^{DR}$
20%	30%	Bias	0.0052	0.2072	-0.0098	0.0019	0.0043
		SSE	0.1175	0.2025	0.1829	0.1846	0.1846
		SEE	0.1761	0.2031	0.1738	0.1842	0.1745
		CP	0.953	0.839	0.938	0.948	0.940
	70%	Bias	-0.0048	0.1826	-0.0425	-0.0078	0.0021
		SSE	0.3031	0.3710	0.3433	0.3557	0.3549
		SEE	0.3038	0.3771	0.3111	0.4125	0.3168
		CP	0.956	0.928	0.929	0.960	0.923
50%	30%	Bias	0.0011	-0.2553	0.0562	0.0042	0.0014
		SSE	0.1759	0.2488	0.2152	0.2185	0.2169
		SEE	0.1758	0.2549	0.2002	0.2721	0.2042
		CP	0.953	0.835	0.920	0.971	0.936
	70%	Bias	-0.0062	-0.1648	0.0388	-0.0025	-0.0020
		SSE	0.3034	0.4381	0.4044	0.4208	0.4171
		SEE	0.3037	0.4408	0.3615	0.5440	0.3695
		CP	0.950	0.934	0.925	0.980	0.925

SSE: Estimated standard error; SEE: Monte Carlo empirical standard error; CP: Empirical coverage probability.

As suggested by a referee, we also conducted an experiment to check the robustness of the parametric and semi-parametric estimators proposed in Section 3. Here, the censoring variable C followed an exponential distribution with rate $\lambda_c e^{0.5Z^3}$ conditional on Z, and λ_c was set to generate censoring rates with 30% and 70%. Note that the true $m(\omega)$ is equal to $1/[1+\lambda_c \exp(-\beta_0 z + 0.5z^3)]$ under the above setting, and we consider the model $m_0(w,\theta) = 1/[1+\exp(-\theta_1 - \theta_2 x - \theta_3 z)]$ to fit $m(\omega)$. All other settings are the same as before; the simulation results are presented in Table 2. As can be seen from the table, the bias for $\widetilde{\beta}_n$ is significantly larger than those of $\widetilde{\beta}_n^{IPW}$ and $\widetilde{\beta}^{DR}$, and the situation is much worse for larger censoring rates and/or missing rates. Thus the augmented inverse probability weighted estimators appear to be quite robust.

5. Data Analysis

We illustrate our methods with a data set from a clinical trial that evaluated tamoxifen as a treatment for stage II breast cancer among elderly women, see Cummings et al. (1986). One hundred seventy elderly women were considered eligible and analyzed in this trial, although we restrict our attention to the 79 women who died by the end of the trial. Among this subset, the censoring rate for the observed individuals and the missing rate are approximately 72% and 23%,

Table 3. Proposed parameter estimates (and jackknife estimates of the standard errors) based on the breast cancer data set.

	Z_1=Treatment	Z_2=Tumor Size	Z_3=Node Count
$\tilde{\beta}_n$	0.0215 (0.016)	-0.0245 (0.017)	-0.0280 (0.016)
$\tilde{\beta}_n^{IPW}$	0.0310 (0.580)	-0.1413 (0.590)	-0.6330 (0.430)
$\tilde{\beta}_n^{DR}$	0.0056 (0.017)	-0.0254 (0.017)	-0.0240 (0.017)

respectively, since 44 women died of breast cancer, 17 died from other known causes, and the cause of death was unknown for 18 women. In this example, X is the survival time (in days)of the individual observation, δ is an indicator showing whether death was due to breast cancer, ξ is an indicator of whether cause of death was known, and $Z = (Z_1, Z_2, Z_3)$ is a vector of covariates, wherefore $W = (X, Z_1, Z_2, Z_3)$ is four-dimensional. Here Z_1 denotes whether the individual received tamoxifen treatment or placebo, Z_2 denotes whether the subject had a tumor size less than 3cm, and Z_3 denotes whether the patient had 4-10 positive axillary lymph nodes.

The estimates of regression coefficients of the three covariates, together with the corresponding jackknife standard error estimators, are presented in Table 3. All three proposed methods led to the same conclusion that none of the covariates appear predictive of the breast-cancer survival. Under the Cox model, time to breast-cancer death was not significantly affected by whether a woman received tamoxifen or placebo, whether her primary tumor was larger or smaller than 3cm, or whether the patient had more or fewer than four positive axillary lymph nodes. Note that the semiparametric augmented inverse probability weighted estimator had the largest SE. This was attributable to the sample size of 79 in the presence of dimension 4.

6. Concluding Remark

When $\pi(\cdot)$ is unknown and there are many covariates involved in the survival data, the 'curse of dimensionality' in fitting multivariate nonparametric regression functions can arise, and a semiparametric approach may be more suitable than fully parametric or fully nonparametric modelling. We might consider the idea of a single index model and introduce an index-variable to fit the conditional probability of missingness, see Ichimura (1993). Let $\pi(X_i, Z_i) = \widetilde{\pi}(X_i, \alpha^\top Z_i)$, where α is a unknown p-dimensional parameter vector and $\tilde{\pi}(\cdot,\cdot)$ is an unknown function. Since ξ is a binary variable, based on the log likelihood

$$\log L(\alpha) = \sum \Big[\xi_i \log \tilde{\pi}_n(X_i, \alpha^\top Z_i) + (1-\xi_i)\log\{1 - \tilde{\pi}_n(X_i, \alpha^\top Z_i)\}\Big],$$

where $\tilde{\pi}_n(\cdot)$ is an nonparametric estimate of $\tilde{\pi}(\cdot)$, the semiparametric maximum likelihood estimator $\hat{\alpha}_n$ for α is a root of $\partial \log L(\alpha)/\partial\alpha = 0$. Letting $(X_i, \widehat{\alpha}_n^\top Z_i)$

be W_i^*, we get a semiparameric estimate $\widetilde{\pi}_n(W^*)$ for $\pi(W_i)$. Klein and Spady (1993) showed that $\hat{\alpha}_n$ is root-n consistent. This method will be discussed in a separate paper.

Appendix A: Proof of Theorem 2.1

Some regularity conditions required for the proofs are listed.

(A1) For each $\theta \neq \theta_0$,

$$\int m(w,\theta)I[m(w,\theta)=0]d\bar{H}_1(w)$$
$$=0=\int(1-m(w,\theta))I[m(w,\theta)=1]d\bar{H}_1(w)$$

and $\int I[m(w,\theta) \neq m(w,\theta_0)]\, d\tilde{H}_1(w) > 0$, where $\bar{H}_1(w) = \mathbb{P}(W \leq w, \xi=1)$.

(A2) $\log m(w,\theta)$ and $\log(1-m(w,\theta))$ are upper-semicontinuous with probability one.

(A3) $\log m(w,\theta)$ and $\log\{1-m(w,\theta)\}$ are uniformly integrable from above under θ_0 for at least a sufficiently small neighborhood $V_\epsilon(\theta^*)$ of θ^*.

(A4) $m(w,\theta)$ possesses continuous partial derivatives of second order with respect to θ at each $\theta \in \Theta$ and the definition region of w; $\nabla m(w,\theta_0)$ is continuous on $[0,\tau]$ and, for each $\theta \in \Theta$, there exists $\varepsilon > 0$ such that $E_{\theta_0}J(W,\theta,\varepsilon) < \infty$, where

$$J(W,\theta,\varepsilon) = \sup\Big\{\Big|\frac{\partial^2 m(w,\theta)}{\partial\varphi_r\partial\varphi_s}\Big| : \|\varphi-\theta\| \leq \varepsilon,\ 1 \leq r,\ s \leq p\Big\}.$$

(A5) $E[D_r m(W,\theta_0)/m(W,\theta_0)]^2 < \infty$ and $E[D_r m(W,\theta_0)/(1-m(W,\theta_0))]^2 < \infty$ for all $1 \leq r \leq p$.

(A6) The matrix $I(\theta_0) = (\sigma_{r,s})_{p\times p} > 0$.

As in the classical approaches of Andersen and Gill (1982) and Fleming and Harrington (1991), a two-step method is employed to show (i) in Theorem 2.1. We first show the result with time truncated, then complete the proof by showing tightness as a second step. Let $\tau_H = \sup\{t : 1-H(t) > 0\}$. For any fixed positive number $\tau < \tau_H$, let

$$\widetilde{U}(\beta,\tau) = \sum\Big\{Z_i - \bar{Z}(\beta,X_i)\Big\}\Big\{\xi_i\delta_i + (1-\xi_i)m_0(W_i,\hat{\theta}_n)\Big\}I(X_i \leq \tau).$$

Note that $\widetilde{U}(\beta,\tau_H) = \widetilde{U}(\beta)$. We first show the asymptotic normality of $\widetilde{U}(\beta_0,\tau)$.

From the proof of Theorem 8.4.1 of Fleming and Harrington (1991), we know that $\int_0^\tau \lambda_0(t)dt < \infty$ and

$$\sup_{0\le x\le\tau,\beta\in\mathcal{B}} \| \bar{Z}(\beta,x) - \bar{z}(\beta,x) \| = o_p(1), \tag{A.1}$$

where $\mathcal{B}$ is any compact neighborhood of β_0 and $\| \cdot \|$ denotes the Euclidean norm of a vector.

To present the asymptotic results of $U(\beta_0,\tau)$, time is truncated and we let

$$\begin{aligned}
\Sigma_1(\tau) &= \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]^{\bigotimes 2} m_0(W_1,\theta_0)I(X_1\le\tau)\Big\},\\
\Sigma_2(\tau) &= \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]^{\bigotimes 2} m_0(W_1,\theta_0)[1-m_0(W_1,\theta_0)][1-\pi(W_1)]I(X_1\le\tau)\Big\},\\
\Sigma_3(\tau) &= \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]\nabla^\top m_0(W_1,\theta_0)I(X_1\le\tau)\Big\},\\
\Sigma_4(\tau) &= \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]\nabla^\top m_0(W_1,\theta_0)\pi(W_1)I(X_1\le\tau)\Big\},\\
\Sigma_0(\tau) &= \Sigma_1(\tau)-\Sigma_2(\tau)+\Sigma_3(\tau)I^{-1}(\theta_0)\Sigma_3^\top(\tau)-\Sigma_4(\tau)I^{-1}(\theta_0)\Sigma_4^\top(\tau).
\end{aligned}$$

Note that, by (C1), (A.1), the Strong Law of Large Numbers and Lenglart's inequality, we can show that $\Sigma_1(\tau) = \int_0^\tau v(\beta_0,s)s^{(0)}(\beta_0,s)\lambda_0(s)ds$; this is the asymptotic variance in Theorem 3.2 of Andersen and Gill (1982). Based on the above notation and the condition

(C2′) The matrix $\Sigma_1(\tau)$ is positive definite,

we state the asymptotic normality of $\widetilde{U}(\beta_0,\tau)$.

Lemma A.1. *Under* (C1) *and* (C2′), *together with* (A1)−(A6), *we have* $n^{-1/2}\widetilde{U}(\beta_0,\tau) \to_d \mathcal{N}(0,\Sigma_0(\tau))$.

Proof. Consider a decomposition of $\widetilde{U}(\beta,\tau)$,

$$\begin{aligned}
\widetilde{U}(\beta,\tau) &= \sum[Z_i-\bar{Z}(\beta,X_i)]\delta_i I(X_i\le\tau)\\
&\quad+\sum[Z_i-\bar{Z}(\beta,X_i)][m_0(W_i,\theta_0)-\delta_i](1-\xi_i)I(X_i\le\tau)\\
&\quad+\sum[Z_i-\bar{Z}(\beta,X_i)][m_0(W_i,\hat{\theta}_n)-m_0(W_i,\theta_0)](1-\xi_i)I(X_i\le\tau)\\
&:= U_1(\beta,\tau)+U_2(\beta,\tau)+U_3(\beta,\tau).
\end{aligned}$$

And let $\widetilde{U}_2(\beta,\tau) = \sum[Z_i-\bar{z}(\beta,X_i)][m_0(W_i,\theta_0)-\delta_i](1-\xi_i)I(X_i\le\tau)$. By (C1), it can be shown that $U_2(\beta_0,\tau) = \widetilde{U}_2(\beta_0,\tau)+o_p(n^{1/2})$. Let

$$\widetilde{U}_3(\beta_0,\tau) = [\Sigma_3(\tau)-\Sigma_4(\tau)]I^{-1}(\theta_0)\sum\frac{\xi_i[\delta_i-m_0(W_i,\theta)]}{m_0(W_i,\theta)[1-m_0(W_i,\theta)]}\nabla m_0(W_i,\theta_0).$$

By (A1)$-$(A6), arguments similar to those in Dikta (1998) can be used to show that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum \frac{\xi_i[\delta_i - m_0(W_i,\theta_0)]}{m_0(W_i,\theta_0)[1-m_0(W_i,\theta_0)]}\nabla m_0(W_i,\theta_0) + o_p(1). \tag{A.2}$$

By (C1) and (A.2), it can be shown that $U_3(\beta_0,\tau) = \widetilde{U}_3(\beta_0,\tau) + o_p(n^{1/2})$. Hence, it is sufficient to derive the asymptotic distribution of $U_1(\beta,\tau) + \widetilde{U}_2(\beta,\tau) + \widetilde{U}_3(\beta,\tau)$.

For $U_1(\beta,\tau)$, by counting processes and martingale theory as in Andersen and Gill (1982),

$$n^{-1/2}U_1(\beta_0,\tau) \longrightarrow_d \mathcal{N}(0,\Sigma_1(\tau)). \tag{A.3}$$

Note that $\widetilde{U}_2(\beta_0,\tau)$ and $\widetilde{U}_3(\beta_0,\tau)$, are sums of i.i.d. variables with mean zero. Hence, by the Central Limit Theorem, we have

$$n^{-1/2}\widetilde{U}_2(\beta_0,\tau) \longrightarrow_d \mathcal{N}(0,\Sigma_2(\tau)), \tag{A.4}$$

$$n^{-1/2}\widetilde{U}_3(\beta_0,\tau) \longrightarrow_d \mathcal{N}\Big(0, [\Sigma_3(\tau)-\Sigma_4(\tau)]I^{-1}(\theta_0)[\Sigma_3^\top(\tau)-\Sigma_4^\top(\tau)]\Big). \tag{A.5}$$

We next consider the covariance structures between the three terms. Note that $\xi_i(1-\xi_i)$ is equal to zero, and then

$$\mathrm{Cov}\Big[n^{-1/2}\widetilde{U}_2(\beta_0,\tau), n^{-1/2}\widetilde{U}_3(\beta_0,\tau)\Big] = 0. \tag{A.6}$$

By (C1) and (A.1), we can show that $\mathbb{E}|\bar{Z}(\beta_0,X_i) - \bar{z}(\beta_0,X_i)| = o(1)$ using Serfling (1980, p.11). Furthermore, the indictors ξ_i and δ_i are independent given X_i and Z_i, and $\mathbb{E}\{\delta_i[m_0(W_j,\theta_0)-\delta_j]|W_l,\ l=1,\ldots,n\} = 0$ for $i \neq j$. We now calculate the other two covariances directly. As $n\to\infty$,

$$\begin{aligned}
&\mathrm{Cov}\Big[n^{-1/2}U_1(\beta_0,\tau), n^{-1/2}\widetilde{U}_2(\beta_0,\tau)]\\
&\quad = n^{-1}\mathbb{E}\Big\{\sum[Z_i - \bar{Z}(\beta_0,X_i)][Z(X_i)-\bar{z}(\beta_0,X_i)]^\top\delta_i[m_0(W_i,\theta_0)-\delta_i]\\
&\qquad \times(1-\xi_i)I(X_i\le\tau)\Big\}\\
&\quad = \mathbb{E}\Big\{[Z_1 - \bar{Z}(\beta_0,X_1)][Z(X_1)-\bar{z}(\beta_0,X_1)]^\top\delta_1[m_0(W_1,\theta_0)-\delta_1]\\
&\qquad \times(1-\xi_1)I(X_1\le\tau)\Big\}\\
&\quad \to \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]^{\bigotimes 2}m_0(W_1,\theta_0)[m_0(W_1,\theta_0)-1][1-\pi(W_1)]I(X_1\le\tau)\Big\}\\
&\quad = -\Sigma_2(\tau),
\end{aligned} \tag{A.7}$$

$$
\begin{aligned}
&\mathrm{Cov}[n^{-1/2}U_1(\beta_0,\tau), n^{-1/2}\widetilde{U}_3(\beta_0,\tau)] \\
&\quad \to \mathbb{E}\Big\{[Z_1-\bar{z}(\beta_0,X_1)]\nabla^\top m_0(W_1,\theta_0)\pi(W_1)\Big\}I^{-1}(\theta_0)[\Sigma_3^\top(\tau)-\Sigma_4^\top(\tau)] \\
&\quad = \Sigma_4(\tau)I^{-1}(\theta_0)[\Sigma_3^\top(\tau)-\Sigma_4^\top(\tau)]. \qquad \text{(A.8)}
\end{aligned}
$$

We complete the proof by combining (A.3)−(A.8).

Proof of Theorem 2.1. The process $\{n^{-1/2}\tilde{U}(\beta_0,\tau)\}$ is tight since $n^{-1/2}\tilde{U}(\beta_0,\tau)$ $-\tilde{U}(\beta_0,\tau_H)) \xrightarrow{p} 0$ and $\lim_{\tau\to\tau_H}\Sigma_1(\tau)=\Sigma_1(\tau)=\Sigma_1(\tau_H)$ and, together with Lemma A.1, it will imply Theorem 2.1 (i).

(ii) By Taylor's expansion, we have that

$$
n^{-1/2}\widetilde{U}(\beta_0) = -n^{-1}\frac{\partial\widetilde{U}(\beta^*)}{\partial\beta}\cdot n^{1/2}(\widetilde{\beta}_n-\beta_0), \qquad \text{(A.9)}
$$

where β^* is a vector between $\widetilde{\beta}_n$ and β_0. Let $V(\beta,t)=S^{(2)}(\beta,t)/S^{(0)}(\beta,t)-\bar{Z}(\beta,t)^{\bigotimes 2}$.For $n^{-1}\partial\widetilde{U}(\beta^*)/\partial\beta$, we consider the same decomposition as in the proof of Lemma A.1:

$$
\begin{aligned}
n^{-1}\frac{\partial\widetilde{U}(\beta)}{\partial\beta} &= -n^{-1}\sum V(\beta,X_i)\delta_i - n^{-1}\sum V(\beta,X_i)[m_0(W_i,\theta_0)-\delta_i](1-\xi_i) \\
&\quad -n^{-1}\sum V(\beta,X_i)[m_0(W_i,\hat{\theta}_n)-m_0(W_i,\theta_0)](1-\xi_i). \qquad \text{(A.10)}
\end{aligned}
$$

Note that, by (C1), the quantity $V(\beta,X_i)$ is uniformly bounded, say by C. Then we can show that

$$
n^{-1}\sup_\beta\Big|\sum V(\beta,X_i)[m_0(W_i,\theta_0)-\delta_i](1-\xi_i)\Big| = o_p(1), \qquad \text{(A.11)}
$$

$$
n^{-1}\sup_\beta\Big|\sum_{i=1}^n V(\beta,X_i)[m_0(W_i,\hat{\theta}_n)-m_0(W_i,\theta_0)](1-\xi_i)\Big| = o_p(1). \qquad \text{(A.12)}
$$

By the same argument as in Andersen and Gill (1982) and Fleming and Harrington (1991), we have

$$
\widetilde{\beta}_n = \beta + o_p(1) \quad\text{and}\quad n^{-1}\frac{\partial\tilde{U}(\beta^*)}{\partial\beta} = -\Sigma_1 + o_p(1). \qquad \text{(A.13)}
$$

Hence, Theorem 2.1 (ii) follows from Theorem 2.1 (i), (A.9)−(A.13) and Slutsky's Theorem.

Appendix B: Proof of Theorem 3.1

The proof is similar to that of Theorem 2.1 in Appendix A, and we only show the case with truncated time. For some $\tau\in(0,\tau_H)$, we define $S(\beta_0,\tau)$ as we did $\tilde{U}(\beta_0,\tau)$ in Appendix A, and consider the decomposition,

$$\begin{aligned}
S(\beta,\tau) =& \sum\{Z_i - \bar{Z}(\beta,X_i)\}[m_0(W_i,\theta_0)-\delta_i]\xi_i\Big\{\frac{1}{\pi(W_i)} - \frac{1}{\pi_n(W_i)}\Big\}I(X_i\le\tau)\\
&+\sum\{Z_i-\bar{Z}(\beta,X_i)\}[m_0(W_i,\hat{\theta}_n)-m_0(W_i,\theta_0)]\Big\{1-\frac{\xi_i}{\pi_n(W_i)}\Big\}I(X_i\le\tau)\\
&+\sum\{Z_i-\bar{Z}(\beta,X_i)\}[m_0(W_i,\theta_0)-\delta_i]\Big\{1-\frac{\xi_i}{\pi(W_i)}\Big\}I(X_i\le\tau)\\
&+\sum\{Z_i-\bar{Z}(\beta,X_i)\}\delta_i I(X_i\le\tau)\\
:=& S_1(\beta,\tau)+S_2(\beta,\tau)+S_3(\beta,\tau)+S_4(\beta,\tau).
\end{aligned}$$

We will show that

$$n^{-1/2}S(\beta_0,\tau) \to_d \mathcal{N}\Big(0,\Sigma_1(\tau)+\Sigma_5(\tau)\Big). \tag{B.14}$$

First by (C1),

$$\begin{aligned}
\mathbb{E}|n^{-1/2}S_1(\beta_0,\tau)|^2 &= n^{-1}\sum\mathbb{E}\Big\{|Z_i-\bar{Z}(\beta,X_i)|^2[m_0(W_i,\theta_0)-\delta_i]^2\\
&\qquad\times\xi_i\Big\{\frac{1}{\pi(W_i)}-\frac{1}{\pi_n(W_i)}\Big\}^2 I(X_i\le\tau)\Big\}\\
&\le K^2\mathbb{E}\Big|\frac{1}{\pi(W_i)}-\frac{1}{\pi_n(W_i)}\Big|^2.
\end{aligned}$$

Note that $0 < \inf_w \pi(w) \le \sup_w \pi(w) \le 1$, together with (3.1) implies that $\mathbb{E}\{\sup_w|\pi_n(w)-\pi(w)|^2\}\to 0$, by Serfling (1980). Thus,

$$n^{-1/2}S_1(\beta_0,\tau) = o_p(1). \tag{B.15}$$

Let

$$\begin{aligned}
\widetilde{S}_{21}(\beta,\tau) &= \sum[Z_i-\bar{z}(\beta,X_i)]\nabla^\top m_0(W_i,\theta_0)\Big\{1-\frac{\xi_i}{\pi(W_i)}\Big\}I(X_i\le\tau),\\
\widetilde{S}_{22}(\beta,\tau) &= \sum[Z_i-\bar{z}(\beta,X_i)]\nabla^\top m_0(W_i,\theta_0)\xi_i\Big\{\frac{1}{\pi(W_i)}-\frac{1}{\pi_n(W_i)}\Big\}I(X_i\le\tau)
\end{aligned}$$

and $\widetilde{S}_2(\beta,\tau) = n^{-1/2}\widetilde{S}_{21}(\beta,\tau)+n^{-1/2}\widetilde{S}_{22}(\beta,\tau)$. It is easy to show that $S_2(\beta_0,\tau) = \widetilde{S}_2(\beta_0,\tau)\cdot\sqrt{n}(\widehat{\theta}_n-\theta_0)+o_p(n^{-1/2})$. Furthermore, by the Law of Large Numbers and (C3), we have that $n^{-1}\widetilde{S}_{21}(\beta_0,\tau) = o_p(1)$, and $n^{-1}\widetilde{S}_{22}(\beta_0,\tau) \le \sup_w|\pi_n(w)-\pi(w)|\cdot n^{-1}\sum K\|\nabla m_0(W_i,\theta_0)\| = o_p(1)$. Hence

$$n^{-1/2}S_2(\beta_0,\tau) = o_p(1). \tag{B.16}$$

For the third term in the decomposition, by (A.1), $S_3(\beta_0,\tau) = \widetilde{S}_3(\beta_0,\tau) + o_p(n^{-1/2})$, where

$$\widetilde{S}_3(\beta_0,\tau) = \sum [Z_i - \bar{z}(\beta_0, X_i)][m_0(W_i,\theta_0) - \delta_i]\Big\{1 - \frac{\xi_i}{\pi(W_i)}\Big\} I(X_i \le \tau).$$

Note that $\widetilde{S}_3(\beta_0,\tau)$ is the sum of i.i.d. variables and then, by the Central Limit Theorem,

$$n^{-1/2}\widetilde{S}_3(\beta_0,\tau) \longrightarrow_d \mathcal{N}(0,\Sigma_5(\tau)). \tag{B.17}$$

Recall that δ_i and ξ_i are independent conditioned on $W_1,\dots,W_n$. Processing as at (A.7) and (A.8), we can obtain that

$$\mathrm{Cov}\Big(n^{-1/2}\widetilde{S}_3(\beta_0,\tau), n^{-1/2}S_4(\beta_0,\tau)\Big) \longrightarrow 0. \tag{B.18}$$

Notice that $S_4(\beta,\tau)$ is exactly the term $U_1(\beta,\tau)$ in Appendix A. Hence, the asymptotic normality in (B.14) follows from (B.15)−(B.18), and Andersen and Gill (1982). Some arguments similar to those used in the proof of Theorem 2.1 can then be used to prove Theorem 3.1(i).

The proof of the Theorem 3.2 is similar to that of Theorem 3.1, except that some regularity conditions for $\pi(\cdot,\gamma)$ are required. The proof is omitted.

Acknowledgement

The authors would like to thank referees for their valuable comments. The research of Qihua Wang was supported by the National Science Fund for Distinguished Young Scholars in China (10725106), the Naqtional Natural Science Foundation of China (10671198), the National Science Fund for Creative Research Groups in China, the research grant from Key Lab of Complex Structure and Data Science and the Research Grants Council of the Hong Kong (HKU 7050/06P).

References

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

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Cummings, F., Gray, R., Davis, T., Tormey, D., Harris, J., Falkson, G. and Arseneau, J. (1986). Tamoxifen versus placebo: double-blind adjuvant trial in elderly women with stage ii breast cancer. *NCI Monogr.* **1**, 119-23.

Devroye, L. P. (1978). The uniform convergence of the nadaraya-watson regression function estimate. *Canad. J. Statist.* **6**, 179-91.

Dikta, G. (1998). On semiparametric random censorship models. *J. Statist. Plann. Inference* **66**, 253-279.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York, 1 edition.

Gao, G. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with misssing cause of failure. *Biometrika* **92**, 875-891.

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

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

Hardle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16**, 1428-49.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *J. Econometrics* **58**, 71-120.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1 edition.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387-421.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York, 1st edition.

Lu, K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* **57**, 1191-1197.

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

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York, 1st edition.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1st edition.

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

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York, 1 edition.

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.

Wang, C. Y., Wang, S., Zhao, L. P. and Ou, S. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. *J. Amer. Statist. Assoc.* **92**, 512-525.

Zhou, X. and Sun, L. (2003). Additive hazards regression with missing censoring information. *Statist. Sinica* **13**, 1237-1257.

Department of Statistics and Actuarial Science, The University of Hong Kong.

E-mail: cliu1221@graduate.hku.hk

School of Mathematics and Statistics, Yunnan University, Yunnan 650091, China.
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China.
Department of Statistics and Actuarial Science, The University of Hong Kong.
E-mail: qhwang@amss.ac.cn

(Received December 2007; accepted April 2009)