1. Introduction

The Cox (1972) proportional hazards model has been widely used to study the effects of covariates on a failure time T (Cox (1975); Tsiatis (1981); Andersen and Gill (1982)). Another commonly-used model in survival analysis is the proportional odds model (Bennett (1983); Dabrowska and Doksum (1988); Murphy et al. (1996); Yang and Prentice (1999)). The two models are special cases of linear transformation models given as follows:

$$S_F(t|Z) = g\{h(t) + \beta^T Z\}, \tag{1.1}$$

where $Z = [z_1, \ldots, z_p]^T$ is the p-dimensional covariate vector, $S_F(t|Z) = P(T > t|Z)$ is the survival function of T given Z, the continuous, strictly decreasing link function $g(.)$ is given or specified up to a finite-dimensional parameter, $h(.)$ is a completely unspecified strictly increasing function, and β is a $p \times 1$ vector of unknown regression coefficients. Further, note that model (1.1) has an equivalent form

$$h(T) = -\beta^T Z + \epsilon,$$

where the distribution of the error ϵ is $P(\epsilon \leq x) = F_\epsilon(x) = 1 - g(x)$. Note that when $g(.) = \exp\{-\exp(.)\}$, (1.1) gives the Cox proportional hazard model, and when $g(.) = 1/\{1+\exp(.)\}$ it corresponds to the proportional odds model.

Typically, the failure time T is subject to right censoring. Let C be the censoring time. We assume conditional independence of T and C given Z. Let $X = \min(T, C)$ and $\delta = I_{[T \leq C]}$, where $I_{[.]}$ is the indicator function. For right-censored data, when $g(.)$ is completely specified, Cheng et al. (1995) proposed a class of estimation procedures for estimating regression parameter β in model (1.1). The method was further developed by Cheng et al. (1997), Fine et al. (1998) and Cai et al. (2000). A key step of their approach is the estimation of survival function for censoring variable by Kaplan-Meier estimator. Its validity relies on the assumption that the censoring variable is independent of the covariates. Under the independence of T and C given Z, using martingale arguments, Chen et al. (2002) proposed an estimation procedure for the analysis of right-censored data. Under i.i.d sampling, Chen et al. (2002) derived the asymptotic variance of the proposed estimators. The specific goal of the paper is to extend the methods of Cheng et al. (1995) and Chen et al. (2002) to the case when the sample has been drawn from a population using a complex design.

In population-based surveys, the sample is usually drawn from a finite survey population via a complex design, such as stratified multi-stage sampling. The design parameters for the survey are often related to the true hazard function, but are not explicitly part of the model being fitted. In this case, the i.i.d. assumption no longer holds. Binder (1992) considered a procedure for fitting proportional hazards models to survey data, which has been implemented in statistical analysis software packages, such as SUDAAN. Lin (2000)

provided a formal justification of Binder's method. Furthermore, he presented an alternative approach which regards the survey population as a random sample from an infinite universe and accounts for this randomness in the statistical inference. Boudreau and Lawless (2006) fitted the proportional hazard model to data from both informative and non-informative designs for stratified clustered super-population. They estimated the baseline cumulative hazard function and proposed another variance estimator. For informative designs, they relied on the results given by Lin (2000). In Section 2, based on the approach of Cheng et al. (1995) and Chen et al. (2002), we propose design-based estimators for regression parameter. Furthermore, similar to the approach of Lin (2000), we regard the survey population as a random sample from an infinite universe and accounts for this randomness in the statistical inference. In Section 3, we report some simulation results.

2. The Proposed Estimators

Let $(T_i, C_i, X_i, \delta_i, Z_i)$, for $i = 1, \ldots, N$, denote the survey population values of (T, C, X, δ, Z). In this article, similar to the approach of Lin (2000), we regard the survey population as a random sample from the joint distribution of (T, C, X, δ, Z) rather than as fixed quantities. This was referred as the superpopulation inference by Lin (2000) as opposed to the finite-population inference of Binder (1992). Since the survey population is from an infinite universe, the population size N can go to infinity, which allow us to make analytic inference about superpoulatoin parameters β and $h(.)$ of model (1.1) by taking into account the sampling of the survey population from the superpopulation as well as that of the survey sample from the survey population.

If all the population values $(T_i, C_i, X_i, \delta_i, Z_i)$ $(i = 1, \ldots, N)$ are available, the parameters β and $h(t)$ can be estimated by solving the following two equations (see Chen et al. (2002)):

$$U_1(\beta, h) = \sum_{i=1}^{N} \int_0^{\tau} Z_i[dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0, \tag{2.1}$$

and

$$U_2(\beta, h) = \sum_{i=1}^{N} [dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0 \quad (t \geq 0), \tag{2.2}$$

where $\Lambda(\cdot)$ is cumulative hazard function of the distribution function $1 - g(.)$, $Y_i(t) = I_{[X_i \geq t]}$, $N_i(t) = I_{[X_i \leq t, \delta_i = 1]}$ and $\tau = \inf\{t : P(X_i > t) = 0\}$.

Suppose that a sample of size n is drawn from the population of N units through a complex design. Let P_i denote the inclusion probability of the i^{th} element of the sample. Let $W_i = 1/P_i$ $(i = 1, \ldots, n)$ denote the the sampling weight of the i^{th} element. Based on (2.1)

and (2.2), we then propose to estimate β and $h(t)$ by solving the following two equations:

$$\hat{U}_1(\beta, h) = \sum_{i=1}^{n} W_i \int_0^{\tau} Z_i[dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0, \tag{2.3}$$

and

$$\hat{U}_2(\beta, h) = \sum_{i=1}^{n} W_i[dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0 \quad (t \geq 0). \tag{2.4}$$

Let $\mathcal{H}$ denote be the collection of all nondecreasing step functions on $[0, \infty)$ with jumps only at observed noncensoring times. We denote by $(\hat{\beta}, \hat{h}(t; \hat{\beta}))$ the solution of (2.3) and (2.4). Note that $\hat{h}(t; \hat{\beta})$ is a step function in t that rises at the distinct jump points of $\{I_{[X_i \leq t, \delta_i = 1]}; i = 1, \ldots, n\}$. For the special case of Cox model, i.e. $g(\cdot) = \exp^{-\exp(\cdot)}$ and $\Lambda(t) = \exp(t)$, it then follows from (2.4) that $d[\exp(h(t))] = \sum_{i=1} W_i dN_i(t) / \sum_{j=1}^{t} W_j Y_j(t) \exp(\beta^T Z_j)$. If we plug this into (2.3), we obtain

$$\hat{U}_1(\beta, h) = \sum_{i=1}^{n} \int_0^{\infty} W_i \left\{ Z_i - \frac{\sum_{j=1}^{n} W_j Z_j Y_j(t) \exp(\beta^T Z_j)}{\sum_{j=1}^{n} W_j Y_j(t) \exp(\beta^T Z_j)} \right\} dN_i(t) = 0,$$

which is the weighted estimating equation proposed by Binder (1992). Equations (2.3) and (2.4) suggest the following iterative algorithms for computing $\hat{\beta}$ and $\hat{h}(t; \hat{\beta})$:

Step 0: Choose an initial value of β, denoted by $\hat{\beta}^{(0)}$.

Step 1: Let $t_1 < t_2 < \cdots < t_{n_d} < \tau$ denote the distinct uncensored points and $W_1^d, \ldots, W_{n_d}^d$ be their corresponding weights. Based on (2.4), we obtain $\hat{h}^{(0)}(t_1; \hat{\beta}^{(0)})$ by solving

$$\sum_{i=1}^{n} W_i Y_i(t_1) \Lambda(\beta^T Z_i + h(t_1)) = W_1^d,$$

with $\beta = \hat{\beta}^{(0)}$. Then, obtain $\hat{h}(t_k)$ for $k = 2, \ldots, n_d$, one-by-one by solving the equation

$$\sum_{i=1}^{n} W_i Y_i(t_k) \Lambda(\beta^T Z_i + h(t_k)) = W_k^d + \sum_{i=1}^{n} W_i Y_i(t_k) \Lambda(\beta^t Z_i + h(t_k -)),$$

with $\beta = \hat{\beta}^{(0)}$, where W_k^d the corresponding weight of t_k .

Step 2: Obtain a new estimate of β by solving (2.3) with $h(t_k) = \hat{h}^{(0)}(t_k; \hat{\beta}^{(0)})$.

Step 3: Set $\hat{\beta}^{(0)}$ to be the estimate obtained in Step 2 and repeat Steps 1 and 2 until prescribed convergence criteria are met.

To facilitate theoretical development, we rewrite (2.3) and (2.4) as

$$\hat{U}_1(\beta, h) = \sum_{i=1}^{N} \frac{\zeta_i}{p_i} \int_0^\tau Z_i[dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0, \tag{2.5}$$

and

$$\hat{U}_2(\beta, h) = \sum_{i=1}^{N} \frac{\zeta_i}{p_i}[dN_i(t) - Y_i(t)d\Lambda(\beta^T Z_i + h(t))] = 0 \ \ (t \geq 0), \tag{2.6}$$

where ζ_i indicates, by the values 1 versus 0, whether or not the i^{th} unit of the survey population is selected into the sample, and p_i is the inclusion probability for the i^{th} unit. Notice that $p_{ij} = P(\zeta_i\zeta_j = 1)$ is the joint inclusion probability for both unit i and j, and ζ_i and ζ_j can be dependent to each other, i.e. $p_{ij} \neq p_i p_j$. It is assumed that $p_i > 0$ for all i. Let $Y_i(t) = I_{[X_i \geq t]}$ and $N_i(t) = I_{[X_i \leq t, \delta_i = 1]}$. Let $\mathcal{F}_t$ denote the filtration generated by

$$\sigma\{Z_i, Y_i(x), \delta_i I_{[X_i \leq t]}, I_{[X_i \leq x]}, x \leq t; i = 1, \ldots, N\},$$

where $\sigma(A)$ denote the complete σ-field generated by A. Let β_0 and $h_0(\cdot)$ denote the true values of β and $h(\cdot)$, respectively. Since $U_1(\beta, h)$ and $U_2(\beta, h)$ are the estimating equations for β_0 calculated from a random sample of size N and $N^{-1}\hat{U}_1(\beta, h)$ and $N^{-1}\hat{U}_2(\beta, h)$ converges to the same limit as $N^{-1}U_1(\beta, h)$ and $N^{-1}U_2(\beta, h)$, respectively, it follows by Proposition of Chen et al. (2002) that the estimators $\hat{\beta}$ and $\hat{h}(\cdot; \hat{\beta})$ of are consistent estimators of β_0 and $h_0(\cdot)$, respectively. Let $M_i(t) = N_i(t) - \int_0^t Y_i(s)d\Lambda(\beta_0^T Z_i + h_0(s))$ $(i = 1, \ldots, N)$. Under model (1.1), since

$$\begin{aligned} E[dN_i(t)|Z_i, \mathcal{F}_{t-}] &= P(t \leq T_i < t + dt, T_i < C_i|Z_i, \mathcal{F}_{t-}) \\ &= Y_i(t)d\Lambda(\beta_0^T Z_i + h_0(t)), \end{aligned}$$

we have

$$E[dM_i(t)|Z_i, \mathcal{F}_{t-}] = E[dN_i(t)|Z_i, \mathcal{F}_{t-}] - Y_i(t)d\Lambda(\beta_0^T Z_i + h_0(t)) = 0.$$

It follows that $M_i(t)$ is a martingale process with respect to $\mathcal{F}_t$, Let $G_i(t)$ $(i = 1, \ldots, N)$ be a $p \times 1$ vector of predictable process with respect to $\mathcal{F}_t$ and define

$$M_G(t) = \sum_{i=1}^{N} \int_0^t G_i(s)dM_i(s)$$

and

$$M_{G,\epsilon}(t) = \sum_{i=1}^{N} \int_0^t G_i(s)I_{[|G_i(s)|>\epsilon]}dM_i(s).$$

Similar to proposition of Chen et al. (2002), under suitable regularity conditions, we can show the following asymptotic theorem.

Theorem 1. Suppose that the regularity conditions (Fleming and Harrington, 1991) for ensuring the central limit theorem for counting process martingales holds, i.e. (i) the predicable process $\langle M_G\rangle(t)\xrightarrow{p}V_G(t)$ for all $t\in[0,\tau]$ as $N\to\infty$, where $V_G(0)=0$ and $V_G(t)-V_G(s)$ is positive semidefinite for all $0\le s\le t\le\tau$, and (ii) $\langle M_{G,,\epsilon}\rangle(t)\xrightarrow{p}0$ for $\epsilon>0$ as $N\to\infty$; (iii) $N^{-1/2}\sum_{i=1}^{N}\int_0^\tau\frac{\zeta_i-p_i}{p_i}Z_i dM_i(t)$ is asymptotically zero-mean normal. Then, we have that $N^{\frac{1}{2}}(\hat\beta-\beta_0)\to N(0,\Sigma_{\hat\beta})$ in distribution, as $N\to\infty$, where $\Sigma_{\hat\beta}$ is given by

$$\Sigma_{\hat\beta}=D^{-1}(\beta_0)[\Sigma_1(\beta_0)+V(\beta_0)]D^{-1}(\beta_0), \tag{2.7}$$

where $D(\beta_0)=\lim_{N\to\infty}\frac{1}{N}\frac{\partial}{\partial\beta}U_1(\beta,\hat h(\cdot,\beta))\big|_{\beta=\beta_0}$. The proof is complete.

The covariance matrix $\Sigma_1(\beta_0)$ can be consistently estimated by

$$\hat\Sigma_1(\hat\beta)=\frac{1}{N}\sum_{i=1}^{N}\bigg[\int_0^\tau\frac{\zeta_i}{p_i}[Z_i-\hat\mu_Z(t;\hat\beta)]^{\otimes2}\lambda(\hat h(t;\hat\beta)+\hat\beta^T)Y_i(t)]d\hat h(t;\hat\beta)\bigg],$$

$$a^{\otimes2}=aa^T,\quad \hat\mu_Z(t;\hat\beta)=\frac{\sum_{i=1}^{N}\frac{\zeta_i}{p_i}[Z_i\lambda(h_P(X_i)+\hat\beta^TZ_i)Y_i(t)\hat B(t;X_i)]}{\sum_{i=1}^{N}\frac{\zeta_i}{p_i}[\lambda(\hat h(t;\hat\beta)+\hat\beta^TZ_i)Y_i(t)]},$$

and

$$\hat B(t,s)=\exp\bigg(\int_s^t\frac{\sum_{i=1}^{N}\frac{\zeta_i}{p_i}[\dot\lambda(\hat h(x;\hat\beta)+\hat\beta^TZ_i)Y_i(x)]}{\sum_{i=1}^{N}\frac{\zeta_i}{p_i}[\lambda(\hat h(x;\hat\beta)+\hat\beta^TZ_i)Y_i(x)]}d\hat h(x;\hat\beta)\bigg).$$

The matrix $D(\beta_0)$ can be consistently estimated by

$$\hat D(\hat\beta)=\frac{1}{N}\frac{\partial}{\partial\beta}U_1(\beta,\hat h(\cdot,\beta))\big|_{\beta=\hat\beta}.$$

Next, let Var_m denote the variance with respect to model, i.e. the variation from one survey population to another, and $\mathrm{Var}_{p|m}$ denote the conditional variance with respect to design given model (i.e. given the observations $X_i,\delta_i,Z_i(i=1,\dots,N)$ from one survey population). Now, $V(\beta_0)=V_1(\beta_0)+V_2(\beta_0)$, where

$$V_1(\beta_0)=\lim_{N\to\infty}N^{-1}\mathrm{Var}_m\big(E_{p|m}\big[\sum_{i=1}^{N}V_i(t;\beta_0)\big]\big)$$

and

$$V_2(\beta_0)=\lim_{N\to\infty}N^{-1}E_m[\mathrm{Var}_{p|m}\big(\sum_{i=1}^{N}V_i(t;\beta_0)\big)].$$

Since $E_{p|m}[V_i(t;\beta_0)] = -\int_0^\tau \mu_Z(t;\beta_0)dM_i(t)$, we have

$$V_1(\beta_0) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\int_0^\tau \mu_Z(t;\beta_0)^{\otimes 2}dM_i(t),$$

which can be consistently estimated by

$$\hat{V}_1(\hat{\beta}) = N^{-1}\sum_{i=1}^{N}\frac{\zeta_i}{p_i}\int_0^\tau \hat{\mu}_Z(t;\hat{\beta})^{\otimes 2}dN_i(t).$$

Next, let $p_{ij} = P(\zeta_i\zeta_j = 1|\mathcal{F})$. Then

$$V_2(\beta_0) = \lim_{N\to\infty} N^{-1}E_m\left[\mathrm{Var}_{p|m}\left(\sum_{i=1}^{N} V_i(t;\beta_0)\right)\right]$$

$$= \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} E_m\left[\frac{p_{ij}-p_ip_j}{p_ip_j}(Z_iM_i(\tau))^{\otimes 2}\right],$$

which can be consistently estimated by

$$\hat{V}_2 = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\zeta_i\zeta_j\frac{p_{ij}-p_ip_j}{p_{ij}p_ip_j}(Z_iN_i(\tau))^{\otimes 2}.$$

Hence, $\Sigma_{\hat{\beta}}$ can be consistently estimated by $\hat{\Sigma}_{\hat{\beta}} = \hat{D}^{-1}(\hat{\beta})[\hat{\Sigma}_1(\hat{\beta}) + \hat{V}_1(\hat{\beta}) + \hat{V}_2]\hat{D}^{-1}(\hat{\beta})$.

Notice that assumption (i) of Theorem 1 requires that the $G_i(s)$ function must be appropriate standardized, assumption (ii) of Theorem 1 is a Lindeberg-type condition which essentially guarantees that he influence of any single process is negligible in the limit, and assumption (iii) of Theorem 1 is needed to permit the application of central limit theorem (C.L.T.) to the normalized Horvitz-Thompson estimator; i.e.,

$N^{-1/2}\sum_{i=1}^{N}\int_0^\tau \frac{\zeta_i-p_i}{p_i}Z_idM_i(t)$ is asymptotically zero-mean normal. As far as we know, there does not exist a general theory on the conditions required for the C.L.T. of Horvitz-Thompson estimator. However, for some particular sampling procedures, conditions for asymptotic normality can be found in the literature. We list some cases as follows. For simple random sampling without replacement conditions required for the C.L.T. to hold are found in Hájek (1960), Hájek (1961) and Scott and Wu (1981); for rejective sampling in Hájek (1964); for random replacement sampling in Rosén (1967); for unequal probability sampling without replacement in Rosén (1972) and Práková (1984); for stratified random sampling in Krewski

and Rao (1981) and Bickel and Freedman (1994); and for two-stage sampling in Ohlsson (1989).

3. Simulation Studies

A simulation study is conducted to investigate the performance of $\hat{\beta}$.

Case 1: Proportional odds model

We generated a population of $N = 5000$ lifetimes T using the proportional odds model with $h(t) = \log(t/10)$ and $\beta_0 = (\beta_{01} = 1, \beta_{02} = 2)^T$. The resulting T has the survivorship function

$$P(T > t|Z_1, Z_2) = \frac{1}{1 + \exp\{\log(t/10) + Z_1 + 2Z_2\}},$$

where Z_1 is an ordinal variable with $P(Z_1 = i) = 0.25$ for $i = 1, 2, 3, 4$, Z_2 is a Bernoulli random variable with probability 0.5 and Z_1 is independent of Z_2. Note that the median of T at baseline $Z = (0, 0)^T$ is 10 in this setting. We generated right censoring variable U was generated from $U(0, \theta)$. The values of θ are set at 0.5, 2 and 8. We suppose that the survey was originally designed to estimate the median of T, i.e. $10\exp(-Z_1 - 2Z_2)$. However, only a related size measure, $\exp(-Z_3 - 2Z_2)$, was available, where $Z_3 = Z_1 + U(0, 3)$ is a continuous random variable and the correlation between Z_3 and Z_1 is equal to 0.79. The population was stratified into four strata by ordering the size measure $\exp(-Z_3 - 2Z_2)$ such that the subpopulation sizes of the four strata are equal to $N_1 = 2000$, $N_2 = 1500$, $N_3 = 1000$, and $N_4 = 500$, respectively. For each stratum, total samples of size $n = 100, 200$ were drawn using simple random sampling without replacement. Hence, the inclusion probability of the elements in the i^{th} stratum is equal to n/N_i. For each replication, we generated $N = 5000$ right-censored observations (X_i, δ_i, Z_i) $(i = 1, \ldots, 5000)$. To obtain the estimator $\hat{\beta}$, the value of τ was set at the largest values of X_i's. The R-code was used to generate the simulations. For each simulated dataset, we obtained the estimators $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$. Repeat this procedure 1000 times. Table 1 shows the simulated biases, simulated standard deviations (std), the estimated standard deviation based on the estimated asymptotic covariance matrix

$\hat{\Sigma}_{\hat{\beta}}$ (denoted by $\text{sd}(\hat{\beta}_i)$). An approximate $1-\alpha$ confidence interval for β_{0i} is also constructed using $\hat{\beta}_i \pm z_{\alpha/2}\text{sd}(\hat{\beta}_i)$, where $z_{\alpha/2}$ is the $\alpha/2$ upper percentile point of the standard normal distribution. Table 1 shows the results of the empirical coverage (denoted by $C(\hat{\beta}_i)$ and $C(\tilde{\beta}_i)$ of $1-\alpha = 0.95$ confidence intervals. Based on the estimator $\hat{h}(t;\hat{\beta})$, we also estimate the baseline survival function $S_0(t) = 1/(1+t/10)$ at $t = 5, 20$, denoted by $\hat{S}_0(t) = \frac{1}{1+\exp\{\hat{h}(t;\hat{\beta})\}}$. Table 2 shows the simulated biases and simulated standard deviations (std) of the estimator $\hat{S}_0(t)$. Tables 1 and 2 also show the proportion of right-censoring (denoted by p_c).

Case 2: Cox model

We generated a population of $N = 1000$ lifetimes T using the Cox odds model with hazard function $h(t|Z_1, Z_2) = h_0(t)e^{\beta_1 Z_1 + \beta_2 Z_2}$, with $h_0(t) = e^t$, $\beta_1 = -9$, $\beta_2 = -5$, Z_1 is an ordinal variable with $P(Z_1 = i) = 0.25$ for $i = 1, 2, 3, 4$, Z_2 is a Bernoulli random variable with probability 0.8 and Z_1 is independent of Z_2. We generated right censoring variable C was generated from exponential distribution with mean μ_c equal to 100 and 50. We suppose that the survey was originally designed to estimate the median of T, i.e. $\log(-\log(0.5)) + 9Z_1 + 5Z_2$. However, only a related size measure, $9Z_3 + 5Z_2$, was available, where $Z_3 = Z_1 + N(0, 0.1)$ is a continuous random variable and the correlation between Z_3 and Z_1 is equal to 0.96. The population was stratified into four strata by ordering the size measure $9Z_3+5Z_2$) such that the subpopulation sizes of the four strata are equal to $N_1 = 100$, $N_2 = 200$, $N_3 = 350$, and $N_4 = 350$, respectively. For each stratum, total samples of size $n = 25, 50, 75$ were drawn using simple random sampling without replacement. Hence, the inclusion probability of the elements in the i^{th} stratum is equal to n/N_i. For each replication, we generated $N = 1000$ right-censored observations (X_i, δ_i, Z_i) $(i = 1, \ldots, 1000)$. The simulation results are reported in Table 3.

Table 1. Simulated biases and std of $\hat{\beta}$

				$\hat{\beta}_1$		
θ	n	p_c	bias	std	sd($\hat{\beta}_1$)	$C(\hat{\beta}_1)$
0.5	100	0.57	0.036	0.305	0.288	0.933
0.5	200	0.57	0.024	0.229	0.217	0.938
2.0	100	0.35	0.027	0.271	0.258	0.934
2.0	200	0.35	0.019	0.194	0.190	0.942
8.0	100	0.18	0.018	0.240	0.234	0.939
8.0	200	0.18	0.012	0.176	0.174	0.946
				$\hat{\beta}_2$		
θ	n	p_c	bias	std	sd($\hat{\beta}_2$)	$C(\hat{\beta}_2)$
0.5	100	0.57	0.043	0.476	0.463	0.932
0.5	200	0.57	0.028	0.353	0.377	0.937
2.0	100	0.35	0.032	0.429	0.418	0.936
2.0	200	0.35	0.019	0.321	0.315	0.941
8.0	100	0.18	0.023	0.386	0.374	0.938
8.0	200	0.18	0.015	0.277	0.286	0.945

Table 2. Simulated biases and std of $\hat{S}_0(t)$

			$\hat{S}_0(5)$		$\hat{S}_0(20)$	
θ	n	p_c	bias	std	bias	std
0.5	100	0.57	0.015	0.087	0.018	0.063
0.5	200	0.57	0.014	0.066	0.014	0.049
2.0	100	0.35	0.015	0.083	0.012	0.056
2.0	200	0.35	0.013	0.061	0.011	0.040
8.0	100	0.18	0.012	0.076	0.011	0.053
8.0	200	0.18	0.007	0.054	0.008	0.034

Table 3. Simulated biases and std of $\hat{\beta}$

				$\hat{\beta}_1$		
μ_c	n	p_c	bias	std	sd($\hat{\beta}_1$)	$C(\hat{\beta}_1)$
50	100	0.50	0.036	0.305	0.288	0.933
50	200	0.50	0.024	0.229	0.217	0.938
50	300	0.50	0.024	0.229	0.217	0.938
100	100	0.20	0.018	0.240	0.234	0.939
100	200	0.20	0.018	0.240	0.234	0.939
100	300	0.20	0.012	0.176	0.174	0.946
				$\hat{\beta}_2$		
μ_c	n	p_c	bias	std	sd($\hat{\beta}_2$)	$C(\hat{\beta}_2)$
50	100	0.50	0.043	0.476	0.463	0.932
50	200	0.50	0.043	0.476	0.463	0.932
50	300	0.50	0.028	0.353	0.377	0.937
100	100	0.20	0.023	0.386	0.374	0.938
100	200	0.20	0.023	0.386	0.374	0.938
100	300	0.20	0.015	0.277	0.286	0.945

Based on the results of Tables 1 and 2, we have the following conclusions:

(i) For the estimation of β_i $(i = 1, 2)$, the standard deviations of $\hat{\beta}_1$ and $\hat{\beta}_2$ increase as the proportion of right-censoring p_c increases. When $n = 100$, the estimated asymptotic standard deviations are smaller than the empirical standard deviations for all the cases considered, which makes the coverage of 95% confidence intervals smaller than the nominal level. However, when $n = 200$ and censoring is light (i.e. $p_c = 0.18$), the coverage of 95% confidence intervals is close to the nominal level.

(ii) For the estimation of baseline $S_0(t)$, the standard deviations of both estimators $S_0(5)$ and $S_0(10)$ increase as the proportion of right-censoring p_c increases. When $n = 200$ and censoring is light (i.e. $p_c = 0.18$) the biases and standard deviations of $S_0(t)$ are small.

4. Conclusion

In this article, we have demonstrated that the approach of Chen et al. (2002) can be used to analyze survey data. Simulation study indicates that the proposed estimator performs adequately with moderate sample size. Using the superpopulation approach of Lin (2000), we can make inferences about parameters which have clear probabilistic interpretations. Further research is required for fitting semiparametric linear model to survey data with left-censored or left-truncated observations.

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.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.*, **2**, 273-277.

Bickel, P. J. and Freedman, D. A. (1994). Asymptotical normality and the bootstrap in stratified sampling. *Ann. Statist.*, **12**, 470-482.

Binder, D. A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrka*, **79**, 139-147.

Boudreau, C. and Lawless, J. (2006). Survival analysis based on the proportional hazards model and survey data. *Canad. J. Statist.*, **34**, 203-216.

Cai, T., Wei, L. J. and Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika*, **87**, 867-878.

Chen, K., Jin, Z and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, **89**, 659-668.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 4, 835-845.

Cheng, S. C., Wei, L. J. and Ying, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *J. Am. Statist. Assoc.*, **92**, 227-235.

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

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

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in the two-sample generalized odds-rate model. *J. Am. Statist. Assoc*, **83**, 744-749.

Fine, J. P., Ying, Z. and Wei, L. J. (1998). On the linear transformation model with censored data. *Biometrika*, **85**, 980-986.

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

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hungar. Acad. Sci.*, **5**, 361-374.

Hájek, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Stat.*, **32**, 506-523.

Hájeck, J. (1964). Asymptotic theory of rejective sampling with varying probabilities form a finite population. *Ann. Math. Statist.*, **35**, 1491-1523.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified sample properties of linearization, jackknife and balanced repeated replication methods. *Ann. Statist.*, **9**, 1010-1019.

Lin, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika*, **87**, 37-47.

Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1996). Maximum likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.*, **92**, 968-976.

Ohlsson, E. (1989). Asymptotic normality for two-stage sampling from a finite population. *Probab. Th. Rel. Fields*, **81**, 341-352.

Práǩová, Z. (1984). On the rate of convergence in Samford-Durbin sampling from a finite population. *Statist. Decisions.*, **2**, 339-350.

Rosén, B. (1967). On the contral limit theorem for sum of independent *r.v.Z. Wahrsch verw.Gebiete.*,**7**,48-72.

Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I and II. *Ann. Math. Statist.*, **43**, 373-397, 748-776.

Scott, A. J. and Wu, C. F. J. (1981). On the asymptotic distribution of ratio and regression estimators. *J. Am. Statist. Assoc.*, **76**, 98-120.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.*, **9**, 93-108.

Yang, S and Prentice, R. (1999). Semiparametric inference in the proportional odds regression model. *J. Am. Statist. Assoc.*, **94**, 125-136.