Semiparametric linear transformation models for interval-censored data

Liuquan Sun

(Academy of Mathematics and Systems Science, CAS)

Jianguo Sun

(University of Missouri-Columbia, USA)

- **Introduction**
- **Models**
- **Estimation Method**
- **Simulation Studies**
- **Data Analysis**
- **Concluding**

1. Introduction

• **Interval-censored data**

Let T be the survival time, L and R are the observation times. Instead of observing T, we only observe L and R such that $L \leq T \leq R$.

For example: In AIDS studies, the HIV infection time and the AIDS incubation time.

• **Current status data (Case I Interval-censored data)**

Let T be the survival time, C the monitoring or observation time. Then all observed information about T is that $T < C$ or $T \geq C$ with C observed.

For example: In carcinogenicity experiments, animals are randomly assigned to various dose of a suspected and are examined at sacrifice or death time for evidence of a malignancy.

Let $F(t) = P(T \leq t)$. A number of authors have studied $F(t)$ based on interval-censored data.

• **Parametric models**: Maximum likelihood estimator.

• **Nonparametric models**: Nonparametric maximum likelihood estimator.

Self-consistency algorithm (Turnbull, 1964);

Convex minorant algorithm (Groeneboom and Wellner, 1992).

- **Semiparametric models**

  Cox model:

  Full likelihood approach (Finkelstein,1986; Huang, 1996).

  Marginal likelihood method (Satten, 1996).

  Imputation approaches (Satten, Datta and Williamson, 1998; Pan, 2000).

  Proportional odds model:

  Sieve method (Rossini and Tsiatis, 1996; Huang and Tsiatis, 1996).

  Conditional likelihood method (Rabinowitz, Betensky and Tsiatis, 2000).

  Additive hazards model

  Estimating equation method (Lin, Oakes and Ying, 1998).

  Efficient score equation (Martinussen and Scheike, 2002).

  Nonparametric maximum likelihood (Ghosh, 2001).

Next, we discuss semiparametric linear transformation models including both Cox hazards model and the proportional odds model as special cases.

2. Models

Let $Z(t)$ be a p-vector of covariates.

$S_Z(t) = P(T \leq t | Z(u), 0 \leq u \leq t)$.

A semiparametric linear transformation model assumes that

$$S_Z(t) \;=\; g(h(t) + \beta' Z(t))\,, \tag{1}$$

where g is a known continuous and strictly decreasing function, and $h(t)$ is an unknown increasing function, β is an unknown vector of regression parameters.

When the covariate $Z(\cdot)$ is time-independent, the model (1) will give the proportional hazards model if $g(t) = \exp\{-\exp(t)\}$ and we will get the proportional odds model if g^{-1} is the -logit function $(-\log\{x/(1-x)\})$.

3. Estimation Method

- **The monitoring time C is independent of T and $Z(\cdot)$**

Let $\{T_i, C_i, Z_i(\cdot)\}$ $(i = 1, ..., n)$ be independent replicates of $\{T, C, Z(\cdot)\}$.

We observe $\{C_i, \delta_i, Z_i(t); t \leq C_i\, , i = 1, ..., n\}$, where $\delta_i = I(C_i \leq T_i)$.

Define $N_i(t) = \delta_i I(C_i \leq t)$, and $N_i^c(t) = I(C_i \leq t)$, and let $\Lambda_c(t)$ denote the cumulative hazard function of C_i's, $Y_i(t) = I(C_i \geq t)$. Define

$$M_i(t; \Lambda_c) = N_i(t) - \int_0^t Y_i(s) g(h_0(s) + \beta_0' Z_i(s))\, d\Lambda_c(s)\ ,$$

where $\{\beta_0, h_0(t)\}$ are the true values of $\{\beta, h(t)\}$, $i = 1, ..., n$.

Then $\{M_i(t; \Lambda_c)\}$ are martingales with respect to the σ-filtration

$$\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s) : s \leq t, i = 1, ..., n\}.$$

For a given β, a estimator for $h_0(t)$ is given by the solution to

$$\sum_{i=1}^{n} \left\{ dN_i(t) - Y_i(t) g(h(t) + \beta' Z_i(t)) d\hat{\Lambda}_c(t) \right\} = 0, \quad 0 \leq t \leq \tau \qquad (2)$$

where τ is a constant such that $P(C \geq \tau) > 0$, and

$$\hat{\Lambda}_c(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i^c(s)}{\sum_{i=1}^{n} Y_i(s)}\ .$$

Denote the estimator of $h_0(t)$ by $\hat{h}_{a0}(t;\beta)$.

For estimating β_0, we propose to use the solution to $U_a(\beta) = 0$, where

$$U_a(\beta) \;=\; \sum_{i=1}^{n}\int_0^{\tau} Q(t)Z_i(t)\left\{dN_i(t) - Y_i(t)g(\hat{h}_{a0}(t;\beta) + \beta' Z_i(t))d\hat{\Lambda}_c(t)\right\}$$

and $Q(t)$ is a known weight process.

$Q(t) = 1$ is the log-rank weight function, and $Q(t) = n^{-1}\sum_{i=1}^{n} Y_i(t)$ is the Gehan weight function.

Let $\hat{\beta}_a$ denote the solution to $U_a(\beta) = 0$ and $\hat{h}_{a0}(t) \equiv \hat{h}_{a0}(t;\hat{\beta}_a)$ the corresponding estimator of the unknown baseline function $h_0(t)$.

We show that both $\hat{\beta}_a$ and $\hat{h}_{a0}(t)$ always exist, are unique and consistent.

We also show that $n^{-1/2}U_a(\beta_0)$ is asymptotically normal with mean zero and covariance matrix that can be consistently estimated by

$$\hat{\Sigma}_a \;=\; \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^{\tau} Q^2(t)\{Z_i(t) - \bar{Z}_a(t)\}^{\otimes 2}dN_i(t) + \int_0^{\tau}\frac{n^2 R_a(t)^{\otimes 2}}{\{\sum_{j=1}^{n} Y_j(t)\}^2}dN_i^c(t)\right.$$

$$\left. - \int_0^{\tau}\frac{nQ(t)\{Z_i(t) - \bar{Z}_a(t)\}R_a'(t)}{\sum_{j=1}^{n} Y_j(t)}dN_i(t) - \int_0^{\tau}\frac{nQ(t)R_a(t)\{Z_i(t) - \bar{Z}_a(t)\}'}{\sum_{j=1}^{n} Y_j(t)}dN_i(t)\right],$$

where $v^{\otimes 2} = vv'$ for a vector v,

$$R_a(t) \;=\; n^{-1}\sum_{i=1}^{n} Q(t)\{Z_i(t) - \bar{Z}_a(t)\}Y_i(t)g(\hat{h}_{a0}(t) + \hat{\beta}_a' Z_i(t))$$

$\dot{g}(t) = dg(t)/dt$, and

$$\bar{Z}_a(t) \;=\; \frac{\sum_{i=1}^{n} Y_i(t)\dot{g}(\hat{h}_{a0}(t) + \hat{\beta}_a' Z_i(t))Z_i(t)}{\sum_{i=1}^{n} Y_i(t)\dot{g}(\hat{h}_{a0}(t) + \hat{\beta}_a' Z_i(t))}\,.$$

Then it follows that $n^{1/2}(\hat{\beta}_a - \beta_0)$ is asymptotically normal with zero mean and covariance matrix that can be consistently estimated by $\hat{A}^{-1}\hat{\Sigma}_a\hat{A}^{-1}$, where

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Q(t)Y_i(t)\{Z_i(t) - \bar{Z}_a(t)\}^{\otimes 2}\dot{g}(\hat{h}_{a0}(t) + \hat{\beta}_a' Z_i(t))d\hat{\Lambda}_c(t)\,.$$

Let $0 = t_1 < t_2 < \cdots < t_K < t_{K+1} = \tau$ denote the set of all jump time points of $\{N_i(t); 0 \leq t \leq \tau, i = 1, ..., n\}$. For the determination of $\hat{\beta}_a$ and $\hat{h}_{a0}(t)$, it is easy to see that the equation (2) and $U_a(\beta) = 0$ can be rewritten as

$$\sum_{i=1}^{n}\left\{\Delta N_i(t_l) - Y_i(t_l)g(h(t_l) + \beta' Z_i(t_l))\Delta\hat{\Lambda}_c(t_l)\right\} = 0,\, l = 1, \ldots, K+1 \qquad (3)$$

and

$$\sum_{i=1}^{n}\sum_{l=1}^{K+1} Q(t_l)Z_i(t_l)\left\{\Delta N_i(t_l) - Y_i(t_l)g(\hat{h}_{a0}(t_l;\beta) + \beta' Z_i(t_l))\Delta\hat{\Lambda}_c(t_l)\right\} = 0\,, \quad (4)$$

where $\Delta N_i(t)$ and $\Delta\hat{\Lambda}_c(t)$ denote the jumps of $N_i(t)$ and $\hat{\Lambda}_c(t)$ at t, respectively.

• **Model Checking**

If model (1) is valid, for any two distinct weight Q_1 and Q_2, the two corresponding estimates $\hat{\beta}_{a1}$ and $\hat{\beta}_{a2}$ of β_0 should be close to each other. On the other hand, if model (1) is incorrect, $\hat{\beta}_{a1}$ and $\hat{\beta}_{a2}$ would differ. This suggests that one can check the adequacy of model (1) by comparing $\hat{\beta}_{a1}$ and $\hat{\beta}_{a2}$. Let $\hat{A}_k$, $\hat{\Sigma}_{ak}$, $\bar{Z}_{ak}$ and R_{ak} be $\hat{A}$, $\hat{\Sigma}_a$, $\bar{Z}_a$ and R_a defined above associated with the weight process Q_k $(k = 1, 2)$. One way for testing the goodness-of-fit of model (1) is to use the statistic

$$S = n(\hat{\beta}_{a1} - \hat{\beta}_{a2})'\hat{\Gamma}^{-1}(\hat{\beta}_{a1} - \hat{\beta}_{a2})\,, \tag{5}$$

which has an asymptotic chi-squared distribution with p degrees of freedom under model (1), where

$$\hat{\Gamma} = \hat{A}_1^{-1}\hat{\Sigma}_{a1}\hat{A}_1^{-1} + \hat{A}_2^{-1}\hat{\Sigma}_{a2}\hat{A}_2^{-1} - \hat{A}_1^{-1}\hat{D}\hat{A}_2^{-1} - \hat{A}_2^{-1}\hat{D}'\hat{A}_1^{-1}\,,$$

with

$$\begin{aligned}\hat{D} \;=\; & n^{-1}\sum_{i=1}^{n}\Bigg[\int_0^\tau Q_1(t)Q_2(t)\{Z_i(t) - \bar{Z}_{a1}(t)\}\{Z_i(t) - \bar{Z}_{a2}(t)\}'dN_i(t) \\ & + \int_0^\tau \frac{n^2 R_{a1}(t)R'_{a2}(t)}{\{\sum_{j=1}^n Y_j(t)\}^2}dN_i^c(t) - \int_0^\tau \frac{nQ_1(t)\{Z_i(t) - \bar{Z}_{a1}(t)\}R'_{a2}(t)}{\sum_{j=1}^n Y_j(t)}dN_i(t) \\ & - \int_0^\tau \frac{nQ_2(t)\{Z_i(t) - \bar{Z}_{a2}(t)\}R'_{a1}(t)}{\sum_{j=1}^n Y_j(t)}dN_i(t)\Bigg]\,.\end{aligned}$$

• **The monitoring time C is dependent of T and $Z(\cdot)$**

Now we consider the situation where T, C and Z may depend on each other, but given Z, we assume that T is independent of C. Also we assume that the hazard function of C at time t has the form

$$d\Lambda_c(t \mid Z) = \exp\{\gamma_0' Z(t)\}\, d\Lambda_{c,0}(t), \tag{6}$$

where $\Lambda_{c,0}(t)$ is an unspecified baseline cumulative hazard function and γ_0 is a vector of unknown regression parameters. Define

$$M_i^*(t; \Lambda_{c,0}) = N_i(t) - \int_0^t Y_i(s) g(h_0(s) + \beta_0' Z_i(s)) \exp\{\gamma_0' Z_i(s)\} d\Lambda_{c,0}(s).$$

Then $\{M_i^*(t; \Lambda_{c,0})\}$ are martingales with respect to the σ-filtration $\mathcal{F}_t$.

For given β and γ, a estimator for $h_0(t)$ is given by the solution to

$$\sum_{i=1}^n \Big[dN_i(t) - Y_i(t) g(h(t) + \beta' Z(t)) \exp\{\gamma' Z_i(t)\} d\hat{\Lambda}_{c,0}(t;\gamma)\Big] = 0, \tag{7}$$

where

$$\hat{\Lambda}_{c,0}(t;\gamma) = \sum_{i=1}^n \int_0^t \frac{dN_i^c(s)}{\sum_{i=1}^n Y_i(s) \exp\{\gamma' Z_i(s)\}}$$

Denote the estimator of $h_0(t)$ by $\hat{h}_{b0}(t; \beta, \gamma)$.

For estimating β_0, we propose to use the solution to $U_b(\beta, \gamma) = 0$ for given γ, where

$$U_b(\beta, \gamma) = \sum_{i=1}^{n} \int_0^{\tau} Q(t) Z_i(t) \left[dN_i(t) - Y_i(t) g(\hat{h}_{b0}(t; \beta, \gamma) + \beta' Z(t)) \exp\{\gamma' Z_i(t)\} d\hat{\Lambda}_{c,0}(t; \gamma) \right].$$

Note that γ is usually unknown. A natural estimate of γ is given by the maximum partial likelihood estimate defined as the solution to

$$U(\gamma) = \sum_{i=1}^{n} \int_0^{\tau} \{Z_i(t) - \bar{Z}(t; \gamma)\} dN_i^c(t) = 0$$

where

$$\bar{Z}(t; \gamma) = \frac{\sum_{i=1}^{n} Y_i(t) \exp\{\gamma' Z_i(t)\} Z_i(t)}{\sum_{i=1}^{n} Y_i(t) \exp\{\gamma' Z_i(t)\}}.$$

Let $\hat{\beta}_b$ and $\hat{\gamma}$ denote the estimators given by $U_b(\beta, \gamma) = 0$ and $U(\gamma) = 0$ and $\hat{h}_{b0}(t) = \hat{h}_{b0}(t; \hat{\beta}_b, \hat{\gamma})$.

We show that both $\hat{\beta}_b$ and $\hat{h}_{b0}(t)$ always exist, are unique and consistent.

It can be shown that $n^{-1/2}\{U_b'(\beta_0,\gamma_0),U'(\gamma_0)\}'$ has asymptotically a normal distribution with mean zero and covariance matrix that can be consistently estimated by $\hat{\Sigma}_b = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{12}' & \hat{\Sigma}_{22} \end{pmatrix}$, where

$$\hat{\Sigma}_{11} = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^\tau Q^2(t)\{Z_i(t)-\bar{Z}_b(t)\}^{\otimes 2}dN_i(t) + \int_0^\tau \frac{n^2 R_b(t)^{\otimes 2}}{\{\sum_{j=1}^n Y_j(t)\exp(\hat{\gamma}'Z_j(t))\}^2}dN_i^c(t)$$

$$-\int_0^\tau \frac{nQ(t)\{Z_i(t)-\bar{Z}_b(t)\}R_b'(t)}{\sum_{j=1}^n Y_j(t)\exp\{\hat{\gamma}'Z_j(t)\}}dN_i(t) - \int_0^\tau \frac{nQ(t)R_b(t)\{Z_i(t)-\bar{Z}_b(t)\}'}{\sum_{j=1}^n Y_j(t)\exp\{\hat{\gamma}'Z_j(t)\}}dN_i(t)\Bigg],$$

$$\hat{\Sigma}_{12} = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\int_0^\tau Q(t)\{Z_i(t)-\bar{Z}_b(t)\}\{Z_i(t)-\bar{Z}(t;\hat{\gamma})\}'dN_i(t)$$

$$-\int_0^\tau \frac{nR_b(t)\{Z_i(t)-\bar{Z}(t;\hat{\gamma})\}'}{\sum_{j=1}^n Y_j(t)\exp\{\hat{\gamma}'Z_j(t)\}}dN_i^c(t)\Bigg],$$

$$\hat{\Sigma}_{22} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{Z_i(t)-\bar{Z}(t;\hat{\gamma})\}^{\otimes 2}dN_i^c(t),$$

with

$$R_b(t) = n^{-1}\sum_{i=1}^{n} Q(t)\{Z_i(t)-\bar{Z}_b(t)\}Y_i(t)g(\hat{h}_{b0}(t)+\hat{\beta}_b'Z_i(t))\exp\{\hat{\gamma}'Z_i(t)\}$$

and

$$\bar{Z}_b(t) = \frac{\sum_{i=1}^n Y_i(t)Z_i(t)\dot{g}(\hat{h}_{b0}(t)+\hat{\beta}_b'Z_i(t))\exp\{\hat{\gamma}'Z_i(t)\}}{\sum_{i=1}^n Y_i(t)\dot{g}(\hat{h}_{b0}(t)+\hat{\beta}_b'Z_i(t))\exp\{\hat{\gamma}'Z_i(t)\}}.$$

It then follows that $n^{1/2}(\hat{\beta}_b - \beta_0)$ is asymptotically normally distributed with mean zero and covariance matrix that can be consistently estimated by

$$(\hat{B}^{-1}, -\hat{B}^{-1}\hat{C})\hat{\Sigma}_b(\hat{B}^{-1}, -\hat{B}^{-1}\hat{C})',$$

where

$$\hat{B} = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Q(t)Y_i(t)\{Z_i(t) - \bar{Z}_b(t)\}^{\otimes 2}\dot{g}(\hat{h}_{b0}(t) + \hat{\beta}_b' Z_i(t))d\hat{\Lambda}_{c,0}(t)$$

and $\hat{C} = \hat{C}_1\hat{C}_2^{-1}$ with

$$\hat{C}_1 = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Q(t)\{Z_i(t) - \bar{Z}_b(t)\}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}' dN_i(t)$$

and

$$\hat{C}_2 = n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) - \bar{Z}(t;\hat{\gamma})\}^{\otimes 2} dN_i^c(t)\,.$$

We also can use two estimators of β_0 corresponding to two different weight processes to check the adequacy of models (1) and (6).

Remark. As shown in (2) and (7), the estimators $\hat{h}_{l0}(t), l = a, b$ may not be monotone in t. To ensure monotonicity, we set $\hat{h}_{l0}^*(t) = \max_{s\le t}\hat{h}_{l0}(s)$. Under appropriate regularity conditions, $\hat{h}_{l0}^*(t)$ and $\hat{h}_{l0}(t)$ are asymptotically equivalent in the sense that $\hat{h}_{l0}^*(t) - \hat{h}_{l0}(t) = o_p(n^{-1/2})$.

4. Simulation Studies

In the study, we considered the linear transformation models with

$$g(t) = \begin{cases} \{1+\alpha\exp(t)\}^{-1/\alpha} & \text{if } \alpha > 0, \\ \exp\{-\exp(t)\} & \text{if } \alpha = 0. \end{cases} \tag{8}$$

The above class of models give the proportional hazards model and the proportional odds model with $\alpha = 0$ and $\alpha = 1$, respectively.

we used $h(t) = \log(0.5t)$ and $\alpha = 0, 0.5$ or 1 and assumed that Z was a Bernoulli random variable with success probability 0.5. The monitoring times were generated from the exponential distribution with hazard rate $\lambda_{c,0}\exp(\gamma_0 Z)$ with $\lambda_{c,0} = 0.5, 1$ or 1.5, and $\gamma_0 = 0$ or 1. The results presented below are based on 2000 samples of size $n = 400$.

Table 1. Simulation results for transformation models with independent monitoring time

			Log-rank				Gehan			
α	β_0	$\lambda_{c,0}$	Mean	SE	SEE	CP	Mean	SE	SEE	CP
0	0.5	0.5	0.5136	0.2633	0.2782	0.9525	0.5146	0.3604	0.3730	0.9530
		1.0	0.5167	0.3226	0.3359	0.9465	0.5241	0.5285	0.5433	0.9455
		1.5	0.5238	0.3813	0.3923	0.9545	0.5342	0.7032	0.7273	0.9550
	1.0	0.5	1.0457	0.3294	0.3390	0.9485	1.0652	0.4499	0.4610	0.9470
		1.0	1.0556	0.3642	0.3667	0.9505	1.0710	0.5635	0.5719	0.9555
		1.5	1.0549	0.4033	0.4110	0.9560	1.0873	0.7491	0.7626	0.9575
0.5	0.5	0.5	0.5178	0.3195	0.3265	0.9545	0.5251	0.4487	0.4542	0.9535
		1.0	0.5227	0.3950	0.3962	0.9460	0.5304	0.6132	0.6321	0.9455
		1.5	0.5253	0.4502	0.4632	0.9555	0.5323	0.8294	0.8488	0.9575
	1.0	0.5	1.0414	0.3359	0.3399	0.9510	1.0503	0.4567	0.4603	0.9515
		1.0	1.0371	0.3828	0.3930	0.9520	1.0652	0.6107	0.6210	0.9560
		1.5	1.0485	0.4338	0.4576	0.9570	1.0716	0.8082	0.8190	0.9600
1.0	0.5	0.5	0.5142	0.3617	0.3793	0.9515	0.5220	0.5057	0.5228	0.9550
		1.0	0.5153	0.4437	0.4631	0.9550	0.5271	0.7227	0.7335	0.9510
		1.5	0.5219	0.5388	0.5544	0.9555	0.5321	0.9722	0.9854	0.9545
	1.0	0.5	1.0341	0.3701	0.3825	0.9475	1.0437	0.5237	0.5489	0.9465
		1.0	1.0561	0.4199	0.4318	0.9525	1.0708	0.6763	0.6989	0.9540
		1.5	1.0517	0.5065	0.5245	0.9565	1.0827	0.8763	0.8950	0.9570

Table 2. Simulation results for transformation models with dependent monitoring time

			Log-rank				Gehan			
α	β_0	$\lambda_{c,0}$	Mean	SE	SEE	CP	Mean	SE	SEE	CP
0	0.5	0.5	0.5102	0.2171	0.2209	0.9510	0.5088	0.3449	0.3557	0.9515
		1.0	0.5096	0.2375	0.2401	0.9485	0.5105	0.4406	0.4556	0.9470
		1.5	0.5108	0.2593	0.2624	0.9550	0.5144	0.5493	0.5661	0.9545
	1.0	0.5	1.0253	0.2384	0.2429	0.9525	1.0213	0.3771	0.3879	0.9525
		1.0	1.0368	0.2610	0.2647	0.9515	1.0381	0.4526	0.4685	0.9530
		1.5	1.0265	0.2778	0.2905	0.9560	1.0284	0.5285	0.5464	0.9555
0.5	0.5	0.5	0.5127	0.2330	0.2426	0.9515	0.5163	0.3827	0.3971	0.9525
		1.0	0.5140	0.2744	0.2751	0.9505	0.5154	0.5013	0.5193	0.9545
		1.5	0.5136	0.2942	0.2994	0.9530	0.5165	0.6008	0.6197	0.9510
	1.0	0.5	1.0184	0.2554	0.2638	0.9525	1.0197	0.4013	0.4232	0.9535
		1.0	1.0269	0.2760	0.2844	0.9470	1.0281	0.4784	0.4861	0.9465
		1.5	1.0297	0.3027	0.3182	0.9545	1.0331	0.5791	0.5983	0.9550
1.0	0.5	0.5	0.5192	0.2683	0.2691	0.9495	0.5198	0.4232	0.4309	0.9475
		1.0	0.5125	0.2982	0.3066	0.9510	0.5145	0.5305	0.5501	0.9525
		1.5	0.5141	0.3273	0.3326	0.9535	0.5180	0.6501	0.6615	0.9545
	1.0	0.5	1.0289	0.2795	0.2844	0.9525	1.0320	0.4418	0.4506	0.9535
		1.0	1.0307	0.3095	0.3127	0.9515	1.0337	0.5132	0.5293	0.9520
		1.5	1.0331	0.3215	0.3286	0.9530	1.0370	0.6281	0.6467	0.9560

5. Data Analysis

In this section, we apply the proposed methodology to the tumorigenicity experiment. The experiment involves a total of 144 RFM mice assigned to either a germ-free or a conventional environment and one of the objectives of the study was to compare the lung tumor incidence rates between the two groups. Since lung tumors are usually regarded as nonlethal and can be observed only at the death of the animals, only current status data are available. In this case, the monitoring time C is the age at death and the censoring indicator δ is 1 if no lung tumor is present at C and 0 otherwise.

Table 3. Ages-at-death of untreated RFM male mice dying with lung tumors

Necropsy findings	Individual ages at death (days)								
	Conventional mice ($Z = 0$)								
Lung tumor	381,	477,	485,	515,	539,	563,	565,	582,	603,
	616,	624,	650,	651,	656,	659,	672,	679,	698,
	702,	709,	723,	731,	775,	779,	795,	811,	838
No Lung tumor	45,	198,	215,	217,	257,	262,	266,	371,	431,
	447,	454,	459,	475,	479,	484,	500,	502,	503,
	505,	508,	516,	531,	541,	553,	556,	570,	572,
	575,	577,	585,	588,	594,	600,	601,	608,	614,
	616,	632,	632,	638,	642,	642,	642,	644,	644,
	647,	647,	653,	659,	660,	662,	663,	667,	667,
	673,	673,	677,	689,	693,	718,	720,	721,	728,
	760,	762,	773,	777,	815,	886			
	Germfree mice ($Z = 1$)								
Lung tumor	546,	609,	692,	692,	710,	752,	753,	781,	782,
	789,	808,	810,	814,	842,	846,	851,	871,	873,
	876,	888,	888,	890,	894,	896,	911,	913,	914,
	914,	916,	921,	921,	926,	936,	945,	1008,	
No Lung tumor	412,	524,	647,	648,	695,	785,	814,	817,	851,
	880,	913,	942,	986					

Define $Z = 1$ for germ-free mice and $Z = 0$ for conventional mice.

Assume that the tumor onset and death times can be described by models (1) and (6), respectively.

Table 4. Estimates of the regression parameter and their estimated standard errors (in bracket) for the tumorigenicity data

Weight	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$
Log-rank	0.5245 (0.2939)	0.8851(0.3580)	1.0430 (0.4332)
Gehan	0.6962 (0.4195)	0.8109(0.4974)	0.9322 (0.5822)

The P-value is 0.07 for $\alpha = 0$ (the proportional hazards model) and suggests that lung tumors seem to occur earlier in the germ-free environment than the conventional environment.

For the group difference under the proportional hazards model, Huang (1996) obtained the MLE of 0.55 with an estimated standard error of 0.29 with a P-value of 0.054, suggesting that the proposed procedure seems effective and accurate. Assuming the additive hazards model, Lin, Oakes and Ying (1998) gave a P-value of 0.08 and similar results.

To check the adequacy of model (1) for the data, assuming the proportional hazards model and the proportional odds model, we obtained $S = 0.6928$ and 0.1589, using the log-rank and Gehan weight. This corresponds to P-values of 0.4 and 0.7, and indicates that both models seem reasonable for the data.

6. Concluding

- An alternative to the proposed approach is the maximum likelihood method.

  Increasing computational complexity.

  No closed form of the asymptotic variance of parameter estimates.

  Asymptotic investigations would be much hard.

- A related approach is to find the semiparametric efficient score function

  Martinussen and Scheike (2002) discussed the analysis of current status data under the additive hazards model. Although the idea is straightforward, its implementation and asymptotic study are not easy for the semiparametric linear transformation models.

- The selection of weight processes.

  It would be useful to investigate how to find a weight process that gives the optimal efficiency of the proposed estimate of regression parameters if it exists.

- Goodness-of-fit method

  It would be helpful to develop methods for selecting the best-fitting model among reasonable models.

- The case of general interval censored data