Journal of Statistical
Theory & Methods
1994, Vol. 5, pp59~66

Estimation on Modified Proportional Hazards Model[1]

Kwang Ho Lee and Mi Sook Lee [2]

ABSTRACT

Heller and Simonoff(1990) compared several methods of estimating the regression coefficient in a modified proportional hazards model, when the response variable is subject to censoring. We give another method of estimating the parameters in the model which also allows the dependent variable to be censored and the error distribution to be unspecified. The proposed method differs from that of Miller(1976) and that of Buckely and James(1979). We also obtain the variance estimator of the coefficient estimator and compare that with the the Buckely-James Variance estimator studied by Hillis(1993).

1. INTRODUCTION

The presence of censored data is a common situation when analyzing data from clinical trials where patients often survive beyond the end of the

[1] Research was supported by Yeungnam University Research Fund., 1993

[2] Department of Statistics, Yeungnam University, Kyongsan, 712-749, Korea

trial period or are lost to follow-up for some reason. Typically, we cann't observe the real survival times $y_1, \ldots, y_n$, and instead observe

$$Z_i = min(y_i, c_i) \quad (i = 1, \ldots, n) \tag{1-1}$$

where $c_1, \ldots, c_n$ are censor values, together with indicator variables

$$\delta_i = \begin{cases} 1 & \text{if} \quad y_i \leq c_i \quad (uncensored) \\ 0 & \text{if} \quad y_i > c_i \quad (censored) \end{cases}$$

The c_i's are not in general assumed to be random variables.

Cox(1972) showed how covariates may be introduced into the nonparametric analysis for such data in the proportional hazards model

$$\lambda(y; x) = \lambda_0(y)e^{-\beta y}$$

where λ is the hazard function for the survival times with covariates x, λ_0 is an unspecified function and β is an unknown coefficient parameter. Methods of analyzing the model have been studied extensively. For example, see Cox(1972), Oakes(1977), Anderson(1982), Halpern(1982), Heller and Simonoff(1990,1992).

We consider here the modified proportional hazards model

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \ldots, n) \tag{1-2}$$

where the ε_i are independently and identically distributed with unspecified distribution F with mean 0 and finite variance σ^2. In fact, this model is linear regression model.

Model (1-2) was studied by many authors including Miller(1976), Buckely and James(1979), Heller and Simonoff(1990,1992), and Hillis(1993). To estimate α and β, they suggested the values a and b which minimize

$$\int \varepsilon^2 d\hat{F}_{a,b}(\varepsilon),$$

where

$$\hat{F}_{a,b} = 1 - \prod_{i; e_{(i)} \leq \varepsilon} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}$$

is the usual Kaplan-Meier estimator of the error distribution F based on the censored and uncensored residuals $e_{i(a,b)} = z_i - a - bx_i$(Kaplan & Meier(1958), Breslow & Crowley(1974), Peterson(1977)).

In this paper, however, we do not use the Kaplan-Meier estimator as the estimator of the error distribution. Instead, we use the quadratic B-spline estimator. Now, we will introduce the smoothing quadratic B-spline estimator which is well described in Klotz(1982).

2. QUADRATIC B-SPLINE ESTIMATOR OF THE SURVIVAL DISTRIBUTION

We consider sorted uncensored residuals $e_{(i)} = y_{(i)} - bx_{(i)}(i = 1, \ldots, k)$ as the knots, where k is the number of uncensored residuals. When the largest residual is censored, the convention is to redefine it as uncensored so that $\hat{F}_b$ will be defined.

Klotz(1982) proposed following hazard rate estimate by approximating using a B-spline function

$$\hat{H}(e) = \sum_{i=1}^{k} \left\{ \frac{\sum_{l \geq i-1} B_{l,3}(e)}{\sum_{j=1}^{n} \sum_{r \geq i-1} B_{r,3}(e_j)} \right\}. \qquad (2-1)$$

For computing, with knots $e_{i-1} < e_i < e_{i+1}$, we have

$$\sum_{l \geq i-1} B_{l,3}(e) = \begin{cases} 0\,, & e \leq e_{i-1} \\ \frac{(e-e_{i-1})^2}{(e_i-e_{i-1})(e_{i+1}-e_{i-1})}\,, & e_{i-1} \leq e < e_i \\ 1 - \frac{(e_{i+1}-e)^2}{(e_{i+1}-e_i)(e_{i+1}-e_{i-1})}\,, & e_i \leq e < e_{i+1} \\ 1, & e \geq e_{i+1}. \end{cases} \qquad (2-2)$$

where $e_{(0)} = 2e_{(1)} - e_{(2)}$ and $e_{(k+1)} = 2e_{(k)} - e_{(k)}$.

He showed that the estimator (2-1) is a non-negative differentiable monotone increasing function of e on the interval $[0, e_{(k)}]$ and compared the performance of the survival estimator obtained by (2-1) and the Kaplan-Meier estimator(1958).

Thus, the estimator of the error distribution was obtained by

$$\hat{F}_b^{sp} = 1 - e^{-\hat{H}(e)} \qquad (2-3)$$

and the estimator $\hat{F}_b^{sp}$ was shown to be consistent to F by Jerome Klotz.

3. PROPOSED ESTIMATORS OF THE PARAMETERS

3.1 Estimator of α and β

Heller and Simonoff(1990) demonstrated the superiority of the Buckely and James methology over other suggested estimator. We will modify the Buckely-James methology and estimate the censored survival time y_i (in case $\delta_i = 0$) by

$$\begin{aligned} \bar{y}_i(b) &= \hat{E}(y_1 | y_i > c_i) \\ &= x_i b + \frac{\int_{c_i - bx_i}^{\infty} e d\hat{F}_b^{sp}(e)}{1 - \hat{F}_b^{sp}(c_i - bx_i)} \end{aligned} \tag{3-1}$$

where $\hat{F}_b^{sp}(e)$ is the spline estimate of F based on the residuals $e_i(b) = y_i - bx_i$ $(i = 1, \ldots, n)$. But Buckely-James estimate the censored survival time y_i based on the Kaplan-Meier estimator of the error distribution function F.

Thus, from the usual least square method, the estimators of β and α are obtained as

$$\begin{aligned} \hat{\beta} &= \frac{\sum^u y_i(x_i - \bar{x}) + \sum^c \bar{y}_i(\hat{\beta})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \text{and} \qquad \hat{\alpha} &= \frac{\sum^u y_i + \sum^c \bar{y}_i(\bar{\beta})}{n} - \hat{\beta}\bar{x} \end{aligned} \tag{3-2}$$

where $\sum^c$ and $\sum^u$ denote summation over the censored values and the uncensored values, respectively.

3.2. Variance estimator of the estimated slope parameter

Hillis(1993) compared the finite sample properties of the variance estimation methods proposed by Buckely and James(1979), Smith(1986), and Weissfeld and Schneider(1987) for a broad range of error and censoring distribution. He showed that for moderate sample sizes Smith's estimator performs the best. The best variance estimator of the Buckley-James type slope

estimator proposed by Smith(1986) is as follows.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \tilde{\sigma}_i^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{p}_i(\hat{\beta}) - \sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} \qquad (3-3)$$

where

$$\tilde{\sigma}_i^2 = \int e^2 d\hat{F}_{\hat{\beta}}(e) - (1-\delta_i)\left[\frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e^2 \, d\hat{F}_{\hat{\beta}}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}(e)} - \left\{\frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e \, d\hat{F}_{\hat{\beta}}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}(e)}\right\}^2\right],$$

and $$\hat{p}_i(\hat{\beta}) = 1 + \hat{\lambda}(c_i - x_i\hat{\beta})\left[c_i - x_i\hat{\beta} - \frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e \, d\hat{F}_{\hat{\beta}}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}(e)}\right]$$

and $\hat{F}_{\hat{\beta}}$ is the usual Kaplan-Meier estimator of the error distribution F based on the censored and uncensored residuals which is introduced in section 1. Now we will modify the estimator $\hat{\sigma}^2$ by replacing the Kaplan-Meier estimator of the error distribution as the proposed estimator based on the B-spline function given in section 3. Thus, if we define $\tilde{v}_i^2$ and $\hat{q}_i(\hat{\beta})$ instead of $\tilde{\sigma}_i^2$ and $\hat{p}_i(\hat{\beta})$ as follows, respectively

$$\tilde{v}_i^2 = \int e^2 d\hat{F}_{\hat{\beta}}^{sp}(e) - (1-\delta_i)\left[\frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e^2 \, d\hat{F}_{\hat{\beta}}^{sp}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}^{sp}(e)} - \left\{\frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e \, d\hat{F}_{\hat{\beta}}^{sp}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}^{sp}(e)}\right\}^2\right],$$

and $$\hat{q}_i(\hat{\beta}) = 1 + \hat{\lambda}(c_i - x_i\hat{\beta})\left[c_i - x_i\hat{\beta} - \frac{\int_{c_i - x_i\hat{\beta}}^{\infty} e \, d\hat{F}_{\hat{\beta}}^{sp}(e)}{\int_{c_i - x_i\hat{\beta}}^{\infty} d\hat{F}_{\hat{\beta}}^{sp}(e)},\right]$$

then the variance estimator of the slope parameter $\hat{\beta}$ as follows;

$$\hat{v}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \tilde{v}_i^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{q}_i(\hat{\beta}) - \sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} \qquad (3-4)$$

The estimate $\hat{\lambda}(u)$ in the above equation is computed using the life-table method applied to the censored and uncensored residuals $e_i = y_i - \hat{\beta}x_i$ ($i =$

$1, \ldots, n)$. . That is, the residuals are groupedinto fixed intervals $[\ g_j, g_{j+1}\ )$, $j = 1, \ldots, s$,which cover the support of the observed residual distribution. For $u \in [g_k, g_{k+1})$,

$$\hat{\lambda}(u) = \frac{deaths(k)}{(g_{k+1} - g_k)\{alive(k) - .5[deaths(k) + censoredD(k)]\}}$$

where $deaths(k)$ is the number of uncensored residuals in $[g_k, g_{k+1})$, $alive$ (k) is the number of residuals greater than g_k, and $censored(k)$ is the number of censored residuals in $[g_k, g_{k+1})$.

4. PERFORMANCES OF THE PROPOSED ESTIMATORS

We investigate the performance of the proposed estimator of the parameters α and β in the modified proportional hazard model(1-2) and variance estimator of the slope estimator $\hat{\beta}$ through the Monte carlo simulation study and the Standford Heart Transplant data.Details on the Data have been presented in Miller(1979).

We will compared the proposed parameter estimators with the estimators proposed by Miller(1976), and Buckley & James(1979). Those estimators are studied detailly by Smith(1986). On the other hand, the proposed variance estimator is compared with the third Buckley-James variance estimator of the slope parameter which are given in Hillis(1993).

Following table shows the estimating values of the parameter α, β, and the variance of slope estimator β for the Heart Transplant Data.

Table 1. comparisions of the estimates from the Stanford Heart Transplant Data

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{v}(\hat{\beta})$
Miller	2.131	0.003	—
Buckley-James	3.582	-0.028	—
Proposed	2.519	-0.009	0.0017

From the table, we know that both estimates α and β have the value between the Miller's and Buckley-James's estimating values. In paticular,

the variance estimate of slope $\hat{\beta}$ is very small. That is, it represent that $\hat{\beta}$ is consist to the slope β.

Finally, we investigate the performance of the proposed slope estimate $\hat{\beta}$ and it's variance estimator $\hat{V}(\hat{\beta})$. From the Monte Carlo Simulation Study, we compare the bias and the variance estimate of the proposed estimator with that of Smith's estimator. Hillis(1993) showed that for moderate sample size, Smith's variance estimator performs the best among the Buckley-James type estimator.

The simulation structure adopted here had n=50, with X taking values -1.96 (0.08) 1.96, β_0=0.4, and β_1=1. The error term was $N(0, 2.1^2)$ and censoring distribution was exponential, with mean $\frac{1}{3}$, corresponding to the existence of 'early censoring' , where many censored values are considerably smaller than the failure times.

To solve the eqation (3-3), we took the initial estimate

$$\hat{\beta}_0 = \frac{\sum^u y_i(x_i - \bar{x}^u)}{\sum^u (x_i - \bar{x}^u)^2} ,$$

iterated 20 times, and took the last value as $\hat{\beta}$.

Table 2. Bias and Variance Estimate of the Slope

	bias($\hat{\beta}$)	$\hat{V}(\hat{\beta})$
Smith's estimate	0.1130	1.0980
Proposed estimate	-0.0002	0.7329

From table 2, we know the fact that the proposed slope estimate perfomes very well in the sense of both bias and variance. Thus, we recommand that you estimate the slope parameter β in the modified proportional hazard model.

REFERENCES

(1) Buckely, J. & James, I. (1979). Linear Regression with Censored Data. *Biometrika* 66, 429-36.

(2) Heller,G. & Simonoff,J.S.(1990). A Comparision of Estimators for Regression with a Censored Response Variable. *Biometrika* 77, 515-20.

(3) Heller,G. & Simonoff,J.S.(1992). Prediction in Censored Survival Data : A Comparision of the Proportional Hazards and Linear Regression Models. *Biometrics* 48, 101-115.
(4) Jerome Klotz(1982). Spline Smooth Estimates of Survival. *IMS, Lecture Notes; Special Topics Meeting on Survival Analysis* 14-25.
(5) Kaplan,E.L. and Meier,p.(1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53, 457-481.
(6) Miller,R.G.(1976). Least Squares Regression with Censored Data. Biometrika 63, 449-64.
(7) Stephen L. Hillis. (1993). A Comparision of Three BuckelyJames Variance Estimators. *Comm. Statist. Simula.*, 22(4), 955-973.