Journal of the Korean Data & Information Science Society
2017, **28**(5), 1167–1177

http://dx.doi.org/10.7465/jkdi.2017.28.5.1167
한국데이터정보과학회지

Censored varying coefficient regression model using Buckley-James method†

Jooyong Shim[1] · Kyungha Seok[2]

[12]Department of Statistics, Inje University

Received 18 July 2017, revised 18 September 2017, accepted 19 September 2017

Abstract

The censored regression using the pseudo-response variable proposed by Buckley and James has been one of the most well-known models. Recently, the varying coefficient regression model has received a great deal of attention as an important tool for modeling. In this paper we propose a censored varying coefficient regression model using Buckley-James method to consider situations where the regression coefficients of the model are not constant but change as the smoothing variables change. By using the formulation of least squares support vector machine (LS-SVM), the coefficient estimators of the proposed model can be easily obtained from simple linear equations. Furthermore, a generalized cross validation function can be easily derived. In this paper, we evaluated the proposed method and demonstrated the adequacy through simulate data sets and real data sets.

Keywords: Censored regression, generalized cross validation function, least squares support vector machine, pseudo-response variable, varying coefficient model.

1. Introduction

In the field of survival data studies, censored regression models are usually estimated using maximum likelihood estimation. Using the pseudo-response variable and the iterative method for censored regression model is proposed by Buckley and James (1979). Zhou (1992) proposed an M-estimators in a linear model of when the responses are subject to right censoring based on the study of Koul *et al*. (1981). Orbe *et al*. (2003) proposed a censored partial regression model to analyse a response variable and the effect of some input variables under censored observations. Their objective function consists of the penalized weighted least squares and an iterative procedure is used for the estimation, and a new bootstrap technique is derived to make inference on the estimators. Hwang *et al*. (2011) proposed a censored regression model based on a penalized regression with L1-penalty.

† This work was supported by the 2016 Inje University research grant.

[1] Adjunct Professor, Institute of Statistical Information, Department of Statistics, Inje University, Kimhae 50834, Korea.

[2] Corresponding author: Professor, Institute of Statistical Information, Department of Statistics, Inje University, Kimhae 50834, Korea. E-mail: statskh@inje.ac.kr

The varying coefficient model proposed Hastie and Tibshirani (1993), which is linear in the input variables but their coefficients are allowed to change with values of other variables called smoothing variables. They mentioned that the dependence of the regression coefficient on smoothing variables implies a special kind of interaction between each input variable and smoothing variables. Several approaches such as the local regression, the kernel smoothing, the polynomial splines and the smoothing splines are used in estimating the coefficient functions (Fan and Zhang, 2008). Shim and Hwang (2015) proposed a varying coefficient least squares support vector regression. For the varying coefficient model researches, Hoover *et al.* (1998), Yang *et al.* (2006), Lee *et al.* (2012) and Ke *et al.* (2016) are good references.

In this paper we propose a varying coefficient regression with censored data, which incorporates a varying coefficient model with the pseudo-response variable defined by Buckley and James (1979). For the estimation of the proposed censored varying coefficient regression the formulation of LS-SVM (Suykens *et al.*, 2001) is applied. We showed that the proposed method provides an easy estimation by using the formulation of LS-SVM. We carried out numerical studies and showed that the proposed method is quite effective in estimating the regression function and varying coefficients in the censored regression model.

The rest of the paper is outlined as follows. Section 2 gives a brief review of the censored regression. Section 3 presents an estimation method for varying coefficient regression with censored data using LS-SVM. Section 4 provides some simulation and real data results that support the adequacy of the methodology to different situations and finally Section 5 presents the conclusions.

2. Censored regression

We set $\boldsymbol{x}_i \in R^d$ be the input variables vector and $t_i \in R$ be the response (survival) variable corresponding to $\boldsymbol{x}_i$, $i = 1, 2, \cdots, n$. Actually the response variable t_i cannot be observed but the observed variables, $y_i = min(t_i, c_i)$ and $\delta_i = I(t_i \le c_i)$, where $I(\cdot)$ is the indicator function and c_i is the censoring variable of the ith observation. Here we assume $t_i \sim iid(F)$ and $c_i \sim iid(S)$, where F and S are unknown distribution function and survival function respectively. Given $\boldsymbol{x}_i$, we set $f(\boldsymbol{x}_i)$ be the regression function of the response variable. Then $f(\boldsymbol{x}_i)$ can be represented as the linear combination of $\boldsymbol{x}_i$ as follows:

$$f(\boldsymbol{x}_i) = \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta}, \quad i = 1, 2, \cdots, n, \tag{2.1}$$

where β_0 is a bias and $\boldsymbol{\beta}$ is a size of $d \times 1$ regression parameters vector. The survival function S of censoring variables c_i's is unknown. Usually the Kaplan-Meier (1958) estimator or its variants are used.

In the censored regression model, the regression function $f(\boldsymbol{x}_i)$ is estimated based on $(\delta_i, y_i, \boldsymbol{x}_i)$, $i = 1, 2, \cdots, n$. Buckley and James (1979) defined the pseudo-response variable

$$y_i^* = y_i\delta_i + E(t_i|t_i > y_i, \boldsymbol{x}_i)(1 - \delta_i) \tag{2.2}$$

and showed $E(y_i^*|\boldsymbol{x}_i) = E(t_i|\boldsymbol{x}_i)$. Since the pseudo-response variable y_i^* cannot be observed it should be substituted with its estimator. Koul *et al.* (1981) defined another pseudo-response variable y_i^*

$$y_i^* = \frac{\delta_i}{S(y_i)} y_i \tag{2.3}$$

and showed that $E(y_i^*|\boldsymbol{x}_i) = E(t_i|\boldsymbol{x}_i)$ and thus y_i^* has the same model as t_i does. The pseudo-response variable y_i^* in (2.3) substituted with its estimator by replacing S with Kaplan-Meier estimator, where roles of survival and censoring variables are reversed. They proposed the least squares regression of $\hat{y}_i^*$ on $\boldsymbol{x}_i$ in the censored regression model. Their estimator is explicitly defined and easily computable. However, Rupert and Miller (1981) proved that $\hat{y}_i^*$ is somewhat peculiar in that it is zero or inflated. The following section introduces the varying coefficient support vector regression model from Hwang and Shim (2015).

3. Varying coefficient regression model with censored data using LS-SVM

Given the input variables vector $\boldsymbol{x}_i = (x_{i1}, \cdot, x_{id})' \in R^d$ and the smoothing variables vector $\boldsymbol{u}_i = (u_{i1}, \cdots, u_{id}) \in R^{d_u}$, the varying coefficient regression function is given by

$$f(\boldsymbol{x}_i, \boldsymbol{u}_i) \ = \ b_0 + \beta_0(\boldsymbol{u}_i) + \sum_{k=1}^{d} x_{ik}\beta_k(\boldsymbol{u}_i). \tag{3.1}$$

Now assume $\beta_k(\boldsymbol{u}_i)$ for $k = 1, \cdots, d$ is nonlinearly related to the smoothing vector $\boldsymbol{u}_i$ such that $\beta_k(\boldsymbol{u}_i) = \boldsymbol{w}_k'\phi(\boldsymbol{u}_i) + b_k$, where $\boldsymbol{w}_k$ is a corresponding weight vector of dimension d_h to $\phi(\boldsymbol{u}_i)$, and $\phi(\cdot) : R^d \rightarrow R_f^d$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. Then the regression function $f(\boldsymbol{x}_i, \boldsymbol{u}_i)$ in (3.1) can be rewritten as

$$f(\boldsymbol{x}_i, \boldsymbol{u}_i) \ = \ b_0 + \boldsymbol{\omega}_0\phi(\boldsymbol{u}_i) + \sum_{k=1}^{d} x_{ik}\boldsymbol{\omega}_k\phi(\boldsymbol{u}_i). \tag{3.2}$$

Using response variable t_i, $y_i = min(t_i,\ c_i)$ and $\delta_i = I(t_i \leq c_i)$, we consider the optimization problem of the varying coefficient model with censored data as

$$minL = \frac{1}{2}||\boldsymbol{\omega}_0||^2 + \frac{1}{2}\sum_{k=1}^{d}||\boldsymbol{\omega}_k||^2 + \frac{\lambda}{2}\sum_{i=1}^{n} e_i^2 \tag{3.3}$$

subject to

$$e_i = y_i^* - b_0 - \boldsymbol{\omega}_0\phi(\boldsymbol{u}_i) - \sum_{k=1}^{d} x_{ik}\boldsymbol{\omega}_k\phi(\boldsymbol{u}_i), \quad i = 1, \cdots, n,$$

where $\lambda > 0$ is a penalty parameter. The estimator of pseudo-response y_i^* is obtained from Buckley and James (1979) as

$$\hat{y}_i^* \;=\; y_i\delta_i + \left(\hat{f}(\boldsymbol{x}_i, \boldsymbol{u}_i) + \frac{\sum_{\hat{e}_k > \hat{e}_i} v_k \hat{e}_k}{1 - \hat{F}(\hat{e}_i)}\right)(1 - \delta_i), \tag{3.4}$$

where $\hat{f}(\boldsymbol{x}_i,\ \boldsymbol{u}_i) \;=\; b_0 + \hat{\beta}_0(\boldsymbol{u}_i) + \sum_{k=1}^d x_{ik}\hat{\beta}_k(\boldsymbol{u}_i)$, $\hat{e}_i = y_i - \hat{f}(\boldsymbol{x}_i,\ \boldsymbol{u}_i)$, $\hat{F}$ is the Kaplan-Meir estimator of the distribution function F of $\hat{e}_i$'s based on $(\delta_1,\ \hat{e}_1), \cdots, (\delta_n,\ \hat{e}_n)$, and $v_1,\ \cdots,\ v_n$ are the jumps of $\hat{F}$. With $\hat{y}_i^*$'s in (3.4) we build a Lagrange function as

$$L = \frac{1}{2}||\boldsymbol{\omega}_0||^2 + \frac{1}{2}\sum_{k=1}^d ||\boldsymbol{\omega}_k||^2 + \frac{\lambda}{2}\sum_{i=1}^n e_i^2 \tag{3.5}$$

$$+ \sum_{i=1}^n \alpha_i(e_i - \hat{y}_i^* + b_0 + \boldsymbol{\omega}_0'\phi(\boldsymbol{u}_i) + \sum_{k=1}^d x_{ik}\boldsymbol{\omega}_k'\phi(\boldsymbol{u}_i)), \tag{3.6}$$

where α_is are the Lagrange multipliers. Taking partial derivatives of the Lagrange function (3.5) with respect to the primal variables $(\boldsymbol{\omega}_k,\ b_0,\ e_i)$ and α_i, we have

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\omega}_0} &= \mathbf{0} \;\rightarrow\; \boldsymbol{\omega}_0 \;=\; \sum_{i=1}^n \phi(\boldsymbol{u}_i)\alpha_i, \\
\frac{\partial L}{\partial \boldsymbol{\omega}_k} &= \mathbf{0} \;\rightarrow\; \boldsymbol{\omega}_k \;=\; \sum_{i=1}^n x_{ik}\phi(\boldsymbol{u}_i)\alpha_i, \quad k = 1, \cdots, d, \\
\frac{\partial L}{\partial b_0} &= 0 \;\rightarrow\; \sum_{i=1}^n \alpha_i = 0, \\
\frac{\partial L}{\partial e_i} &= 0 \;\rightarrow\; \lambda e_i - \alpha_i = 0, \quad i = 1, \cdots, n, \\
\hat{y}_i^* - b_0 - \boldsymbol{\omega}_0\phi(\boldsymbol{u}_i) &- \sum_{k=0}^{d_x} x_{ik}\boldsymbol{\omega}_k'\phi(\boldsymbol{u}_i) - \frac{1}{\lambda}\alpha_i = 0.
\end{aligned}$$

After eliminating e_is and $\boldsymbol{\omega}_k$s, we have the optimal values of α_is are as follows;

$$\begin{bmatrix} K + (\boldsymbol{x}\boldsymbol{x}') \odot K + \frac{1}{\lambda}\boldsymbol{I} & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b_0 \end{bmatrix} \;=\; \begin{bmatrix} \hat{\boldsymbol{y}}^* \\ 0 \end{bmatrix}, \tag{3.7}$$

where $K = (K_{ij})_{n\times n}$ is an $n \times n$ kernel matrix with (i,j)th element $k(\boldsymbol{u}_i, \boldsymbol{u}_j)$ with a kernel function k, $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$ is an $n \times n$ matrix, $\hat{\boldsymbol{y}}^* = (\hat{y}_1^*, \cdots, \hat{y}_n^*)'$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)'$, and "$\odot$" denotes a component-wise product.

From the equations (3.2) and (3.6), the estimators of $\beta_0(\boldsymbol{u}_t)$ and $\beta_k(\boldsymbol{u}_t)$ are obtained as follows:

$$\begin{aligned}\hat{\beta}_0(\boldsymbol{u}_t) &= \sum_{i=1}^{n} k(\boldsymbol{u}_t, \boldsymbol{u}_i)\hat{\alpha}_i, \\ \hat{\beta}_k(\boldsymbol{u}_t) &= \sum_{i=1}^{n} x_{ik} k(\boldsymbol{u}_t, \boldsymbol{u}_i)\hat{\alpha}_i, \quad k = 0, \cdots, d_z,\end{aligned} \tag{3.8}$$

which leads the predicted regression function given $(\boldsymbol{x}_t,\ \boldsymbol{u}_t)$ as

$$\hat{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) = \hat{\beta}_0(\boldsymbol{u}_t) + \sum_{k=1}^{d} x_{tk}\hat{\beta}_k(\boldsymbol{u}_t). \tag{3.9}$$

From the equations (3.7) $\sim$ (3.9) the estimated regression function given $(\boldsymbol{x}_t, \boldsymbol{u}_t)$ can be represented as the linear combination of $\hat{\boldsymbol{y}}^*$ as follows:

$$\hat{f}(\boldsymbol{x}_t, \boldsymbol{u}_t) = H(\boldsymbol{x}_t, \boldsymbol{u}_t)\hat{\boldsymbol{y}}^*,$$

where $H(\boldsymbol{x}_t, \boldsymbol{u}_t) = (k(\boldsymbol{u}_t, \boldsymbol{u}) + (\boldsymbol{x}_t'\boldsymbol{x}') \odot k(\boldsymbol{u}_t, \boldsymbol{u}), 1)H_0$ and

$$H_0 = \begin{pmatrix} \Omega - \Omega\mathbf{1}(\mathbf{1}'\Omega\mathbf{1})^{-1}\mathbf{1}'\Omega \\ (\mathbf{1}'\Omega\mathbf{1})^{-1}\mathbf{1}'\Omega \end{pmatrix}, \quad \Omega = K + (\boldsymbol{x}\boldsymbol{x}') \otimes K + \frac{1}{\lambda}\boldsymbol{I},$$

where $\mathbf{1} = (1 \cdots 1)'$ is an $n \times 1$ vector with element value 1.

Furthermore, the estimators of $\beta_0(\boldsymbol{u}_t)$ and $\beta_k(\boldsymbol{u}_t)$ are can be represented as the linear combination of $\hat{\boldsymbol{y}}^*$ as follows:

$$\hat{\beta}_0(\boldsymbol{u}_t) = S_0(\boldsymbol{u}_t)\hat{\boldsymbol{y}}^*, \quad \hat{\beta}_k(\boldsymbol{u}_t) = S_k(\boldsymbol{u}_t)\hat{\boldsymbol{y}}^*, \quad k = 1, \cdots, d, \tag{3.10}$$

where $S_0(\boldsymbol{u}_t) = k(\boldsymbol{u}_t, \boldsymbol{u})H_1$, $S_k(\boldsymbol{u}_t) = (\boldsymbol{x}'_{\cdot k} \odot k(\boldsymbol{u}_t, \boldsymbol{u}))H_1$, $\boldsymbol{x}_{\cdot k}$ is kth column of $\boldsymbol{x}$, H_1 is an $n \times n$ matrix with the first n rows of H_0.

From (3.10) we can estimate the variance of $\hat{\beta}_k(\boldsymbol{u}_t)$'s as follows:

$$\hat{Var}(\hat{\beta}_k(\boldsymbol{u}_t)) = \hat{\sigma}^2 S_k(\boldsymbol{u}_t)S_k(\boldsymbol{u}_t), \quad k = 0, \cdots, d,$$

where $\hat{\sigma}^2$ is an estimator of variance of $\hat{y}_i^*$ and we estimate it by $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i^* - \hat{f}_i)^2$.

Thus an approximate $(1-\alpha)100\%$ pointwise confidence interval of $\beta_k(\boldsymbol{u}_t)$ is obtained as follows:

$$\hat{\beta}_k(\boldsymbol{u}_t) \pm \Phi(1-\alpha/2)\hat{Var}(\hat{\beta}_k(\boldsymbol{u}_t))^{1/2},$$

where $\Phi(1-\alpha/2)$ is a $(1-\alpha/2)$th quantile of the standard normal distribution.

The estimator of b_0 also can be expressed as the linear combination of $\hat{\boldsymbol{y}}^*$ as follows:

$$\hat{b}_0 = H_2 \hat{\boldsymbol{y}}^*, \tag{3.11}$$

where H_2 is an $1 \times n$ matrix with the $(n+1)$th row of H_0.

From (3.11) we can estimate the variance of $\hat{\beta}_k(\boldsymbol{u}_t)$s as follows:

$$\hat{Var}(\hat{b}_0) = \hat{\sigma}^2 H_2 H_2'.$$

Since $\hat{y}_i^*$ is a function of $(b_0, \beta_0, \beta_k(\boldsymbol{u}_i))$ as shown in (3.4), we need an iterative procedure to estimate the regression parameters $(b_0, \beta_0, \beta_k(\boldsymbol{u}_i))$ as follows:

(i) Initially set all δ_i's ones $(y_i = \hat{y}_i^*)$, we obtain $\boldsymbol{\alpha}$ and b_0 from (3.7).
(ii) Obtain $\hat{y}_i^*$'s from (3.4).
(iii) Obtain $\boldsymbol{\alpha}$ and b_0 from (3.7).
(iv) Iterate (ii) and (iii) until convergence, $\frac{1}{n}||\hat{f}^{(t+1)}(\boldsymbol{x},\boldsymbol{u}) - \hat{f}^{(t)}(\boldsymbol{x},\boldsymbol{u})||^2 < 1e-6$.

The performance of censored varying coefficient regression is affected by the hyper-parameters such as penalty parameter and the kernel parameters. To select the hyper-parameters we define the cross validation function as follows:

$$CV(\theta) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i^* - \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)^{(-i)})^2,$$

where $\hat{y}_i^*$ is the final estimator of y_i^*, θ is the set of hyper-parameters and $\hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)^{(-i)} = \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i|\theta)^{(-i)}$ is the leave one out regression function estimator. Since for each candidates of parameters, $\hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)^{(-i)}$for $i = 1, \cdots, n$, should be computed, to select parameters using CV function is a computationally burdensome task. By using leaving-out-one lemma by Craven and Wahba (1979), $\hat{y}_i^* - \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)^{(-i)}$ can be approximated to

$$(\hat{y}_i^* - \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i))/(1 - \partial\hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)/\partial\hat{y}_i^*) = (\hat{y}_i^* - \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i))/(1 - H_{ii}),$$

where $H_{ii} = \partial\hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i)/\partial y_i$ is a diagonal element of hat matrix H corresponding to $\hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i) = H_{i.}\boldsymbol{y}$ where $H_{i.}$ is the ith row of H. Substituting H_{ii} by their average $tr(H)/n$, the GCV function can be obtained as

$$GCV(\theta) = \frac{1}{n}\frac{\sum_{i=1}^{n}(\hat{y}_i^* - \hat{f}(\boldsymbol{x}_i,\boldsymbol{u}_i))^2}{(1 - tr(H)/n)^2}.$$

4. Numerical studies

In this section we proceed to numerically study the estimation performance of the proposed method for the censored regression function estimation through the simulated data sets as well as real data sets. In the simulation examples and the small cell lung cancer data set (Ying et al., 1995), we used the Gaussian kernel. The linear kernel is utilized in Stanford heart transplant data (Miller and Halpern, 1982).

4.1. Simulation examples

We carried out experiments to show the performance of the proposed method by comparing with the weighted censored regression (Zhou, 1992) and the censored partial regression (Orbe *et al.*, 2003).

To compare, we generate 50 data sets, in each data set survival times are generated as the follows:

$$
\begin{aligned}
t_i &= f(\boldsymbol{x}_i, u_i) + e_i \\
&= b_0 + \beta_0(u_i) + \beta_1(u_i)x_{i1} + \beta_2(u_i)x_{i2} + e_i,
\end{aligned}
$$

where $u_1 \sim iid\, U(0,1)$, $x_{ik} \sim iidN(0,1)$, $b_0 = 0$, $\beta_0(u_i) = 2cos(2\pi u_i)$, $\beta_1(u_i) = 2sin(2\pi u_i)$ and $\beta_2(u_i) = exp(2u_i - 1)$. Censoring time $c_i \sim iid\, U(0,10)$ and we observe $y_i = min(c_i, t_i)$. The average of 50 censoring proportions is 0.1116.

In the weighted censored regression (Zhou, 1992), the survival time t_i is expressed as

$$t_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 u_i + e_i.$$

In the censored partial regression (Orbe *et al.*, 2003), the survival time t_i is expressed as

$$t_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + h(u_i) + e_i,$$

where $h(u_i)$ is a nonlinear function of u_i and is estimated by splines.

Mean squared error is used for the performance metric,

$$MSE = \frac{1}{100}\sum_{i=1}^{100}(f_i - \hat{f}_i)^2.$$

Averages of 50 MSEs and their standard errors of the proposed method, Zhou (1992) and Orbe *et al.* (2003) are obtained as (0.2046, 0.0141), (4.4658, 0.0649), and (2.7464, 0.0654), respectively. Results show that the proposed method is best in the estimation performance in this example.

For the second simulation example, we generate 50 data sets, survival times in each data set are generated as the follows:

$$t_i = f(\boldsymbol{x}_i, u_i) + e_i = b_0 + \beta_0 + \beta_1 x_{i1} + \beta_2(u_i)x_{i2} + e_i,$$

where $u_1 \sim iid\ U(0,1)$, $x_{ik} \sim iidN(0,1)$, $b_0 = 0$, $\beta_1 = 0.1$ and $\beta_2(u_i) = 2sin(2\pi u_i)$. Censoring times $c_i \sim iid\ U(0,10)$ and we observe $y_i = min(c_i, t_i)$. The average of 50 censoring proportions is 0.1378. In the weighted censored regression (Zhou, 1992), the survival time t_i can be expressed as

$$t_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 u_i + e_i.$$

In the censored partial regression (Orbe *et al.*, 2003), the survival time t_i can be expressed as

$$t_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + h(u_i) + e_i.$$

Averages of 50 MSEs and their standard errors by the proposed method, Zhou (1992) and Orbe *et al.* (2003) are obtained as (0.1290, 0.0125), (0.4524, 0.0151), and (0.4487, 0.0150), respectively, showing that the proposed method is best in the estimation performance in this example.

4.2. Real examples

For the real examples, we use the Stanford heart transplant data (Miller and Halpern, 1982). In this study 55 patients were censored among 157 patients who had complete records. In the linear censored regression model the base 10 logarithm of the survival time is expressed as follows:

$$log_{10}(t_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i,\ i = 1, \cdots, 157,$$

where x_i is the age at the first transplant.

The estimators of $(\beta_0, \beta_1, \beta_2)$ are obtained as (1.353, 0.1069, -0.00167) by Buckley and James (1979), (0.8187, 0.1237, -0.001793) by Zhou (1992). Miller and Halpern (1982) concluded that Buckley and James estimators are not dependent on the particular censoring patterns and reliable estimators in this example. is standardized for the better estimation in the censored partial regression (Orbe *et al.*, 2003) and proposed method such as $x_i^* = \frac{x_i - min(x_i)}{max(x_i) - min(x_i)}$.

In the censored partial regression (Orbe *et al.*, 2003), $log_{10}(t_i)$ can be expressed as follows:

$$log_{10}(t_i) = b_0 + \beta x_i^* + h(x_i^*) + e_i.$$

Take the age as the input variable and smoothing variable, we express $log_{10}(t_i)$ using the proposed method as follows:

$$\begin{aligned} log_{10}(t_i) =& b_0 + \beta_0(x_i^*) + \beta_1(x_i^*) x_i^* + e_i \\ =& b + \omega_0 x_i + \omega_1 x_i^2 + e_i. \end{aligned}$$

The estimators of (b_0, β) are obtained as (-0.3071, 0.4534) by Orbe *et al.*, (2003) and the estimators of (b, ω_0, ω_1) are obtained as (1.3005, 0.0972, -0.00148) by the proposed method. The estimated varying coefficients $\beta_0(x_i^*)$s and $\beta_1(x_i^*)$s are described in Figure 4.1 (left) and (middle), which implies that the survival time seems to be affected positively by the age and affected negatively by the squared age. The estimated regression functions are described in Figure 4.1 (right), which show that estimates by Buckley and James (1979) and the proposed method show the similar survival patterns on the age.

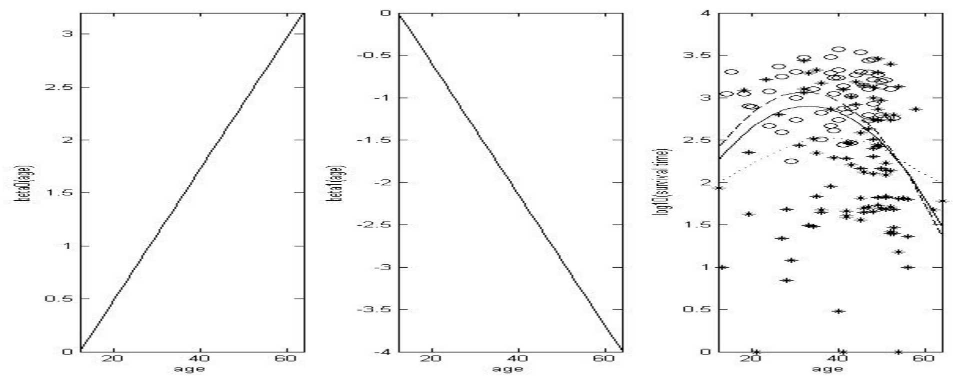

Figure 4.1 The estimated varying coefficients $\beta_0(x)$'s (left) and $\beta_1(x)$'s (middle) by the proposed method. The estimated regression functions superimposed on the scatter plots of the base 10 logarithm of the survival time versus the age (right). o = censored data point, * = uncensored data point. Solid lines are by the proposed method, dashed lines by Buckley and James (1979), dotted lines by Orbe *et al.* (2003)

For the second real example, we use the small cell lung cancer data in Ying *et al.* (1995). In this study, 121 patients are randomly assigned to be treated with either arm A (Cisplatin followed by Etoposide) or arm B (Etoposide followed by Cisplatin). 62 patients are assigned to arm A and 59 to arm B. Ying *et al.* (1995) considered the linear censored regression model as follows:

$$log_{10}(t_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i,\ i = 1, \cdots, 121,$$

where $x_{i1} = 0$ if the ith patient was treated with arm A and 1 if with arm B. $x_{i2} = 0$ is the entry age of the ith patient. The estimators of $(\beta_0, \beta_1, \beta_2)$ are obtained as (2.8666, -0.1253, -0.0023) by Zhou (1992), and (3.028, -0.163, -0.004) by Ying *et al.* (1995). The fact that estimators of β_1 and β_2 are negative indicates that arm A is better than arm B, and survival time decreases as the age at entry increases. But we noticed that the absolute values of estimators are very small.

Using the proposed method, $log_{10}(t_i)$ can be expressed as follows:

$$log_{10}(t_i) = b_0 + \beta_0(u_i) + \beta_1(u_i)x_{i1} + e_i,$$

where $u_i = x_{i2}$ (age) is a smoothing variable.

The estimated regression functions by the proposed method are described in Figure 4.2 (left) and (middle), which show that survival time decreases as the age at entry increases. From Figure 5.2 (right) we can see that $\hat{\beta}_1(u_i) < 0$ for $i = 1, \cdots, 121$, which implies the proposed method also indicate that arm A is better than arm B. The coefficient estimator $\hat{\beta}_1(u_i)$ decrease as age increases indicates that survival time decrease more rapidly at large value of entry than small value.

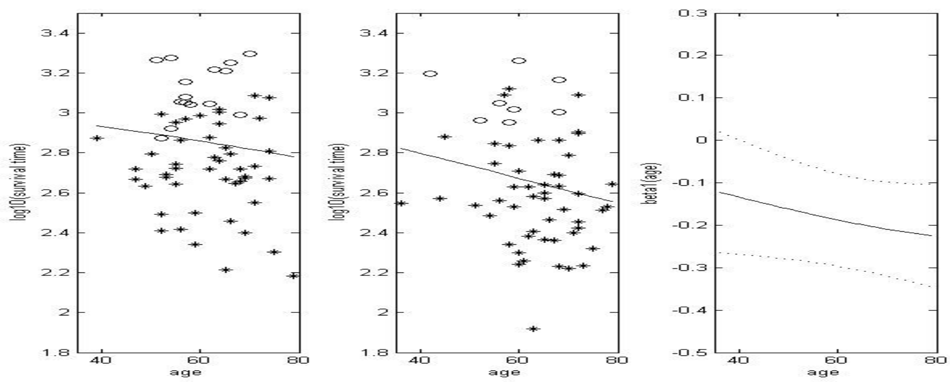

Figure 4.2 The estimated regression functions by the proposed method superimposed on the scatter plots of data points of small lung cancer data for arm A (left) and arm B (middle), o = censored data point, * = uncensored data point. The estimated varying coefficient $\hat{\beta}_1(u_i)$ (right) by the proposed method and their approximate 95% pointwise confidence limits

5. Conclusions

We have proposed a varying coefficient regression with censored data. We showed that the proposed method provides easy estimation by using the formulation of LS-SVM. Numerical studies indicate that the proposed method is quite effective in estimating the regression function and varying coefficients in the censored regression model. Also we showed that the proposed method is simple and reliable.

References

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics*, **31**, 377-403.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, **1**, 179-195.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B*, **55**, 757-796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.

Hwang, C., Kim, M. and Shim, J. (2011). Variable selection in L1 penalized censored regression. *Journal of the Korean Data & Information Science Society*, **22**, 951-959.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, **53**, 457-481.

Ke, Y., Fu, B. and Zhang, W. (2016). Semivarying coefficient multinomial logistic regression for disease progression risk prediction. *Statistics in Medicine*, **35**, 4764-4778.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *The Annal of Statistics*, **9**, 1276-1288.

Lee, Y. K., Mammen, E. and Park, B. U. (2012). Flexible generalized varying coefficient regression models. *Annals of Statistics*, **40**, 1906-1933.

Miller, R. G. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **74**, 301-309.

Orbe, J., Ferreia, E. and Nunes-Anton, V. (2003). Censored partial regression. *Biostatistics*, **4**, 109-121.

Rupert, G. and Miller, Jr. (1981). *Survival analysis*, Wiley, New York.

Shim, J. and Hwang, C. (2015). Varying coefficient modeling via least squares support vector regression. *Neurocomputing*, **161**, 254-259.

Suykens, J. A. K., Vandewalle, J. and De Moor, B. (2001). Optimal control by least squares support vector machines. *Neural Networks*, **14**, 23-35.

Yang, L., Park, B. U., Xue, L. and Hrdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, **101**, 1212-1227.

Ying, Z. L., Jung, S. H. and Wei, L, J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, **90**, 178-184.

Zhou, M. (1992). M-estimation in censored linear models. *Biomerika*, **79**, 837-841.