Journal of the Korean Data & Information Science Society
2011, **22**(5), 951–959

한국데이터정보과학회지

Variable selection in L1 penalized censored regression†

Changha Hwang[1] · Mal Suk Kim[2] · Jooyong Shim[3]

[1]Department of Statistics, Dankook University
[2]Division of Computer Technology, Yeungnam College of Science & Technology
[3]Department of Data Science, Inje University

Received 11 July 2011, revised 29 July 2011, accepted 10 August 2011

Abstract

The proposed method is based on a penalized censored regression model with L1-penalty. We use the iteratively reweighted least squares procedure to solve L1 penalized log likelihood function of censored regression model. It provide the efficient computation of regression parameters including variable selection and leads to the generalized cross validation function for the model selection. Numerical results are then presented to indicate the performance of the proposed method.

Keywords: Censored regression model, generalized cross validation function, iteratively reweighted least squares procedure, L1-penalty, variable selection.

1. Introduction

The censored regression model and the least squares method to accommodate the censored data seem appealing since they are familiar and well understood. Koul *et al.* (1981) gave a simple least squares type estimation procedure in the censored regression model with the weighted observations and also showed the consistency and asymptotic normality of the estimator. Zhou (1992) proposed an M-estimator of the regression parameter of censored regression model based on the weights Koul *et al.* (1981) proposed. Orbe *et al.* (2003) proposed the estimation procedure of censored regression model where estimators of regression parameters and nonlinear function are obtained by minimizing the penalized weighted least squares objective function through iterative method. They also proposed the procedure to generate the bootstrap resamples to obtain the uncertainty measures of estimators. Jin *et al.* (2003) proposed the estimation procedure where regression parameter estimates of censored regression model are obtained from non-monotone estimating equations based on the

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0028135).
[1] Professor, Department of Statistics, Dankook University, 126, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do 448-701, Korea.
[2] Adjunct professor, Division of Computer Technology, Yeungnam College of Science & Technology, Daegu 705-703, Korea.
[3] Corresponding author: Adjunct professor, Department of Data Science, Inje University, Obang-Dong, Kimhae 621-749, Korea. E-mail: ds1631@hanmail.net

weighted log-rank statistics. The estimating equations are solved through iterative method with Gehan (1965)-type estimate as the initial value. Ghosh and Ghosal (2006) proposed the estimation procedure based on a nonparametric Bayesian approach which uses a Dirichlet prior for the mixture of Weibull distribution in the censored regression model. Markov Chain Monte Carlo method is used to obtain the marginal posterior distribution of regression parameters.

In this paper we set $\boldsymbol{x}_i$ be the covariate vector and t_i be the response variables (survival times) corresponding to covariate vector, $\boldsymbol{x}_i$ or transformation on it, where $i = 1, 2, \cdots, n$. In fact we cannot observe t_i's but the observed variable, $y_i = \min(t_i, c_i)$ and $\delta_i = I(t_i \leq c_i)$, where $I(\cdot)$ denotes the indicator function and c_i is the censoring variable corresponding to $\boldsymbol{x}_i$ for $i = 1, 2, \cdots, n$. c_i's are assumed to be independently distributed with unknown survival distribution functions. Let $m(\boldsymbol{x}_i)$ be the regression function of the response variable given $\boldsymbol{x}_i$. We assume that $\boldsymbol{x}_i$'s and y_i's are centered so that $m(\boldsymbol{x}_i)$ is related to the covariate vector $\boldsymbol{x}_i$ in a linear form without a bias as

$$m(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}, \ i = 1, 2, \cdots, n, \tag{1.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ regression parameter vector. Generally, all the p covariates may not much affect the survival times so that some β's may be zeros in true regression function. Many variable selection techniques for linear regression models have been extended to the context of survival models, including the best-subset selection, stepwise selection, and Bootstrap procedures (Sauerbrei and Schumacher, 1992). Recently the LASSO (least absolute shrinkage and selection operator; Tibshirani, 1997) has been proposed for Cox proportional hazards model (Cox, 1972). By shrinking some regression parameters to zero, this method provides the selection of important variables and the estimation of regression parameters simultaneously. Huang *et al.* (2005) proposed the regularization and variable selection approach using LASSO (Tibshirani, 1996). Hu and Rao (2010) proposed a weighted least squares method with censoring constraints and sparse penalization to fit censored regression models with high-dimensional covariates.

We consider the penalized censored regression with L1 norm which is known to have the sparsity on estimation of regression parameters (Williams, 1995). We use the iteratively reweighted least squares (IRWLS) procedure to solve the penalized log likelihood function with L1 norm of censored regression. It provide the efficient computation including variable selection and provides the generalized cross validation function for the model selection.

The rest of paper is organized as follows. In Section 2 we briefly review the censored regression. In Section 3 we propose IRWLS procedure to penalized estimation for censored regression with L1 norm. In Section 4 we perform the numerical studies with simulated data sets and a real data set. In Section 5 we give the conclusions.

2. Censored regression

In most practical cases survival distribution function of c_i's, G, is not known and needs to be estimated by the Kaplan-Meier (1958) estimator or its variation. The problem considered here is that of the estimation of $m(\boldsymbol{x}_i)$ based on $(\delta_1, y_1, \boldsymbol{x}_1), \cdots, (\delta_n, y_n, \boldsymbol{x}_n)$. Buckley and James (1979) defined the pseudo-response variable

$$y_i^* = y_i\delta_i + E(t_i|t_i > y_i, \boldsymbol{x}_i)(1 - \delta_i). \tag{2.1}$$

They showed $E(y_i^*|\boldsymbol{x}_i) = E(t_i|\boldsymbol{x}_i)$ and proposed the iteration method to estimate the regression parameters $\boldsymbol{\beta}$. Koul *et al.* (1981) defined new observable responses y_i^* as $y_i^* = w_i y_i$ with

$$w_i = \frac{\delta_i}{G(y_i)}, \tag{2.2}$$

and showed y_i^* has the same mean as t_i and thus follows the same linear model as t_i does. Here, $\widehat{G}$, the Kaplan-Meier estimates (Kaplan and Meier, 1958) of survival distribution function G of c_i's can be obtained as,

$$\widehat{G}(y) = \begin{cases} \prod_{i:y_{(i)} \le y} \left(\dfrac{n-i}{n-i+1} \right)^{1-\delta_{(i)}}, & \text{if } y \le y_{(n)} \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

where $(y_{(i)}, \delta_{(i)})$ is (y_i, δ_i) ordered on y_i for $i = 1, \cdots, n$. Koul *et al.* (1981) proposed the ordinary least squares regression of y_i^* on $\boldsymbol{x}_i$. Zhou (1992) proposed the weighted least squares regression of y_i^* on $\boldsymbol{x}_i$ with w_i as follows:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{x}'W\boldsymbol{x})^{-1}\boldsymbol{x}'W\boldsymbol{y}, \tag{2.4}$$

where $\boldsymbol{x}$ is a $n \times p$ matrix and W is a diagonal matrix of w_i's. $\widehat{\boldsymbol{\beta}}$ in (2.4) can be seen as the minimizer of the objective function as follows:

$$\frac{1}{2}\sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2. \tag{2.5}$$

3. L1 penalized censored regression

From (2.5) we assume that $r_i = \sqrt{w_i}(y_i - m(\boldsymbol{x_i}))$ follows a probability distribution such that $p(r_i) \propto \exp(-0.5 r_i^2)$. Then the negative log-likelihood of the given data set can be expressed as (constant terms are omitted),

$$\ell(\boldsymbol{m}|\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{n} w_i (y_i - m(\boldsymbol{x}_i))^2. \tag{3.1}$$

The regression function is estimated by a linear model, $m(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$. Then the maximum likelihood estimates of $\boldsymbol{\beta}$ are obtained by minimizing the negative log-likelihood function,

$$\ell(\boldsymbol{\beta}|\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{n} w_i (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2. \tag{3.2}$$

The maximum likelihood estimates of $\boldsymbol{\beta}$ generally lead severe overfitting, we are encouraged to use a prior over $\boldsymbol{\beta}$. Then the penalized maximum likelihood estimates (the maximum a posteriori estimates) of $\boldsymbol{\beta}$ are obtained by minimizing the objective function,

$$L(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}), \tag{3.3}$$

where $p(\boldsymbol{\beta})$ is some prior over $\boldsymbol{\beta}$.

To have the sparsity on estimation of $\boldsymbol{\beta}$, we use a Laplacian prior (Williams, 1995),

$$p(\boldsymbol{\beta}) \propto \exp(-\lambda||\boldsymbol{\beta}||_1), \tag{3.4}$$

where $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_i|$ denotes L1 norm and λ is a positive penalty parameter. The objective can be rewritten as

$$L(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x_i}'\boldsymbol{\beta})^2 + \lambda||\boldsymbol{\beta}||_1. \tag{3.5}$$

Here λ controls the tradeoff between the goodness-of-fit on the data and $||\boldsymbol{\beta}||_1$. The objective function $L(\boldsymbol{\beta})$ in (3.5) is not differentiable with respect to $\boldsymbol{\beta}$, we need a modification of $L(\boldsymbol{\beta})$ for IRWLS procedure.

We define an objective function given $\boldsymbol{\beta}^*$ as

$$L(\boldsymbol{\beta}|\boldsymbol{\beta}^*) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \frac{\lambda}{2}\sum_{j=1}^{p}(\frac{\beta_j^2}{|\beta_j^*|} + |\beta_j^*|), \tag{3.6}$$

then $L(\boldsymbol{\beta}|\boldsymbol{\beta}^*) \geq L(\boldsymbol{\beta})$ with equality if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ (Krishnapuram *et al.*, 2005) and $L(\boldsymbol{\beta}|\boldsymbol{\beta}^*)$ is differentiable with respect to $\boldsymbol{\beta}$. At t th iteration of IRWLS procedure, we have

$$L(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \frac{1}{2}\sum_{i=1}^{n} w_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \frac{\lambda}{2}\sum_{j=1}^{p}(\frac{\beta_j^2}{|\beta_j^*|} + |\beta_j^*|). \tag{3.7}$$

Then $\boldsymbol{\beta}^{(t+1)}$ is obtained by minimizing $L(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)})$ with respect to $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta}^{(t+1)} = (\boldsymbol{x}'W\boldsymbol{x} + \lambda V(\boldsymbol{\beta}^{(t)}))^{-1}\boldsymbol{x}'W\boldsymbol{y}, \tag{3.8}$$

where $V(\boldsymbol{\beta}^{(t)})$ is the diagonal matrix consisted of $1/|\beta_j^{(t)}|, j = 1, \cdots, p$.

During iteration, we find that some β_j's tend to zero keeping the value of objective function $L(\boldsymbol{\beta})$ decreasing. This motivates that we can find sparse estimates of $\boldsymbol{\beta}$ which provides decreasing value of the objective function $L(\boldsymbol{\beta})$ simultaneously. Algorithm of L1 penalized censored regression using IRWLS Procedure is given as follows:

1. Set $v = (1:p)'$ and $\boldsymbol{\beta}(v)^{(0)}$.
2. Find solution $\boldsymbol{\beta}(v)^{(t+1)}$ which minimizes $L(\boldsymbol{\beta}(v)|\boldsymbol{\beta}(v)^{(t)})$.
3. Set $\beta_i^{(t+1)} = 0$ which is sufficiently close to zero. Find $v = \{j|\beta_j^{(t+1)} \neq 0\}$.
4. Iterate 2-3 until $|L(\boldsymbol{\beta}(v)^{(t+1)}) - L(\boldsymbol{\beta}(v)^{(t)})| <$ tolerance.

The functional structures of L1 penalized censored regression is characterized by penalty parameter λ. To select the optimal penalty parameter we define the cross validation (CV) function for the model selection criterion as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n} w_i(y_i - \widehat{m}_\lambda^{(-i)}(\boldsymbol{x}_i))^2, \tag{3.9}$$

which is similar to the weighted residual sum of squares (Zhou, 1998). Here $\widehat{m}_\lambda^{(-i)}(\boldsymbol{x}_i)$ is the regression function estimated without i th observation. Since for each candidate of penalty, $\widehat{m}_\lambda^{(-i)}(\boldsymbol{x}_i)$ for $i = 1, \cdots, n$, should be evaluated, selecting penalty parameter using CV function is computationally burdensome. GCV function is obtained as follows:

$$GCV(\lambda) = \frac{n \sum_{i=1}^{n} w_i (y_i - \widehat{m}_\lambda(\boldsymbol{x}_i))^2}{(n - tr(H))^2}, \tag{3.10}$$

where $H = (\boldsymbol{x}(:, v)' W \boldsymbol{x}(:, v) + \lambda V(v, v))^{-1} \boldsymbol{x}(:, v)' W$ is the hat matrix such that $\widehat{m}_\lambda(\boldsymbol{x}) = H\boldsymbol{y}$ with the (i, j) th element $h_{ij} = \partial \widehat{m}(\boldsymbol{x}_i)/\partial y_j$. Details of derivation of GCV function can be seen in Cho *et al.* (2010), Hwang and Shim (2010), Shim (2005), Shim and Lee (2009). Akaike (1974) defined Akaike's Information Criterion (AIC) for model selection criterion a follows:

$$AIC = 2l(\boldsymbol{\beta}|\boldsymbol{x}) + 2K, \tag{3.11}$$

where K is the number of estimable parameters in the model and $l(\boldsymbol{\beta}|\boldsymbol{x})$ is the negative log-likelihood. Inspired by AIC, we define an AIC-type criterion to incorporatethe simplicity of the model into the model selection criterion as follows:

$$GCV(\lambda)_{AIC} = \log(GCV(\lambda)) + K, \tag{3.12}$$

where K is the number of variables with non-zero regression parameters.

4. Numerical studies

We illustrate the performance of the proposed method for the estimation and the variable selection through the simulated data sets and the real data set.

Example 4.1 We generate 100 data sets to compare the performance of variable selection with the exhaustive search using the weighted least squares regression of Zhou (1992) in (2.4). For each $i = 1, \cdots, 100$, $x_{i1}, \cdots, x_{i6}$ are generated from a uniform distribution, $U(0, 1)$, respectively and (t, c)'s are generated as follows:

$$t_i = 1 + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_{t_i},\ c_i = 1.4 + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_{c_i},\ i = 1, \cdots, 100,$$

where $\boldsymbol{\beta}' = (2, 0, 2, 0, 0, 0)'$, ϵ_{t_i}'s and ϵ_{c_i}'s are generated from normal distributions, $N(0, 0.1)$, respectively. For each data set, the proposed method is applied with the optimal value of the penalty parameter chosen from GCV function in (3.12). The box plots of each $\widehat{\beta_j}$'s by the proposed method and the weighted least squares regression of Zhou (1992) are shown as in Figure 4.1 (Left) and Figure 4.1 (Right), respectively. Also we obtained the mean squared error of $(\widehat{m}(\boldsymbol{x}) - m(\boldsymbol{x}))$ and its standard deviation in each data set. As results we obtained the averages of 100 mean squared errors and their standard deviations for the proposed method as (0.0066, 0.0045) and (0.0106, 0.0059) for the weighted least squares regression of Zhou (1992), respectively, this implies that the proposed method has better estimation performance than the weighted least squares regression of Zhou (1992) in this example. For the exhaustive search using the weighted least squares regression of Zhou

(1992), we use $(6,2)' = 15$ sets of two variables shown in the first column heading of Table 4.1. We divided each dataset into 15 sub-data sets according to 15 sets of two variables such as $\{y_i, \delta_i, x_{i1}, x_{i2}\}_{i=1}^{100}$, $\{y_i, \delta_i, x_{i1}, x_{i3}\}_{i=1}^{100}, \cdots$, $\{y_i, \delta_i, x_{i5}, x_{i6}\}_{i=1}^{100}$, and obtained 15 sets of $(\widehat{\beta}_i, \widehat{\beta}_j)$ and 15 likelihoods $l(\boldsymbol{\beta})$'s in (3.2) for 15 sub-data sets. We computed the averages of 100 $(\widehat{\beta}_i, \widehat{\beta}_j)$'s and 100 $l(\boldsymbol{\beta})$'s for each set of two variables to choose two most important variable, which are shown in the second and the third column headings of Table 4.1. From Figure 4.1 (Left) and Table 4.1 we can see that the proposed method agree with the exhaustive search using the weighted least squares regression of Zhou (1992) in (2.4) for the variable selection of (x_1, x_3) as the most important set of two variables.

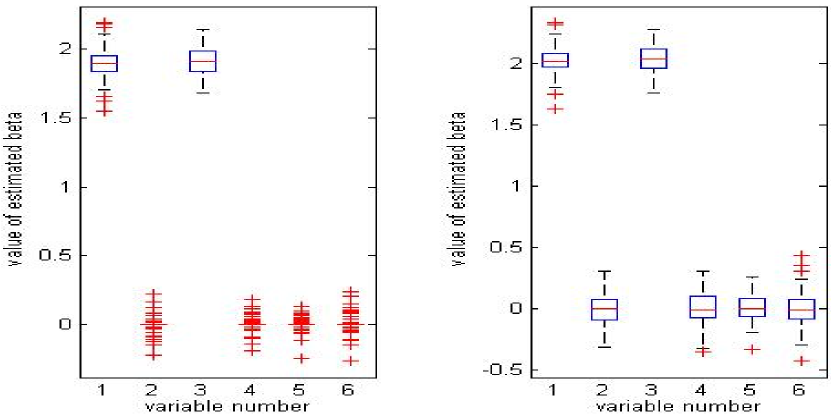

Figure 4.1 Box plots of estimated regression parameters in Example 4.1

Table 4.1 Results of Example 4.1 by exhaustive search (standard deviation in parenthesis)

variables (i, j)	averages of $(\widehat{\beta}_i, \widehat{\beta}_j)$	average of $l(\boldsymbol{\beta})$
(1, 2)	2.1324 (0.2411), 0.0035 (0.2649)	20.8209 (2.6232)
(1, 3)	2.0203 (0.1093), 2.0311 (0.1093)	4.0321 (0.7009)
(1, 4)	2.1219 (0.2433), 0.0181 (0.3058)	20.7286 (2.7369)
(1, 5)	2.1275 (0.2447), 0.0106 (0.2614)	20.8233 (2.6389)
(1, 6)	2.1316 (0.2423), 0.0218 (0.2686)	20.8008 (2.6983)
(2, 3)	0.0100 (0.2669), 2.1437 (0.2326)	20.4846 (2.7192)
(2, 4)	0.0142 (0.3484), -0.0056 (0.4311)	38.9792 (5.7702)
(2, 5)	0.0231 (0.3337), 0.0026 (0.3818)	39.1437 (5.5621)
(2, 6)	0.0226 (0.3518), 0.0202 (0.3839)	39.1222 (5.5674)
(3, 4)	2.1339 (0.2379), -0.0134 (0.2743)	20.4628 (2.6955)
(3, 5)	2.1394 (0.2349), 0.0101 (0.2835)	20.4439 (2.6288)
(3, 6)	2.1396 (0.2382), 0.0039 (0.2740)	20.4625 (2.6787)
(4, 5)	-0.0116 (0.4307), -0.0020 (0.3889)	38.8444 (5.7454)
(4, 6)	-0.0163 (0.4331), 0.0211 (0.3821)	38.8551 (5.7454)
(5, 6)	0.0017 (0.3947), 0.0232 (0.3822)	38.9817 (5.5523)

Example 4.2 We applied the proposed method to the data set of Diffuse large B-cell lymphoma (DLBCL) survival times and gene expression published by Rosenwald *et al.* (2002). This data set consists of expression values of 7399 genes across 240 patients with

DLBCL, including 138 patient deaths during the follow-ups with a median death time of 2.8 years. The patients were divided into a training data of 160 patients and a test data of 80 patients as Bair and Tibshirani (2004). The optimal value of the penalty parameter was chosen as 0.5 from GCV function in (3.12). To evaluate the performance of estimation and variable selection methods, Li (2006), Hu and Rao (2010) proposed to partition the subjects into a high risk and a low risk group according to the estimated survival times, choose the estimated median as the cut-point and then compare the difference between survival times of two risk groups. If there is a significant difference between two risk groups, the selected variables are expected to be highly predictive. Figure 4.2 shows Kaplan-Meier estimates (Kaplan and Meier, 1958) of survival functions for two risk groups of training data (Left) and test data (Right), where the estimated median used as the cut-point of each risk group was obtained by the proposed method. Hu and Rao (2010) selected 79 genes, log-rank tests show highly significant difference between the two groups in survival for training data ($\chi^2_{(1)}$ = 116.12 and p-value < 1e - 10), and for test data ($\chi^2_{(1)}$ = 4.55 and p-value = 0.017). By the gradient LASSO (Huang *et al.*, 2005) 37 genes were selected and resulted in p-value of the log-rank test of test data = 0.05 (Li, 2006). By the proposed method 99 genes were selected, log-rank tests show ($\chi^2_{(1)}$ = 32.17 and p-value < 1e - 9) for training data and ($\chi^2_{(1)}$ = 12.96 and p-value =0.00031) for test data, which show highly significant difference in survival times between the two risk groups for training data and test data. Thus results of log-rank tests indicate better predictive performance of the proposed method than the other methods for test data. Figure 4.3 shows the scatter plots of estimated log-survival times by the proposed method versus log-observed survival times of each risk group in training data (Left) and test data (Right), which are superimposed lines of slope 1, where '*' represents uncensored data point and 'o' represents censored data point.

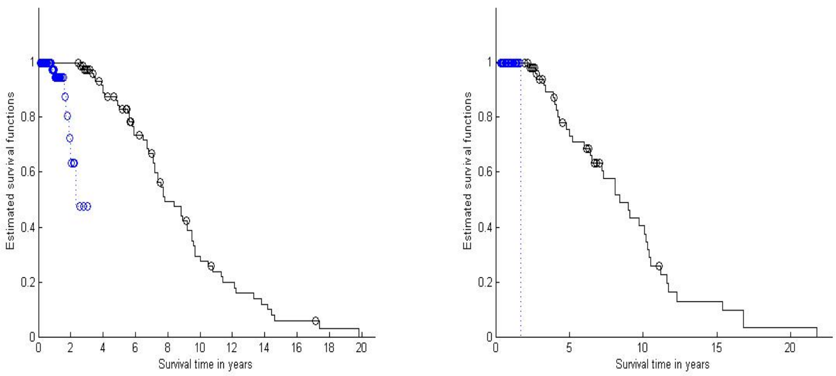

Figure 4.2 Estimated survival functions for two risk groups

5. Conclusions

In this paper, we dealt with estimating the regression function and variable selection based on a penalized censored regression model with L1-penalty. We use the iteratively reweighted least squares procedure to solve L1 penalized log-likelihood function of censored regression model. It provide the efficient computation and leads to the generalized cross validation function for the model selection. Through the examples we showed that the proposed method derives the satisfying solutions. The proposed method is simple and reliable in the point that estimation of the regression function and variable selection are performed simultaneously.

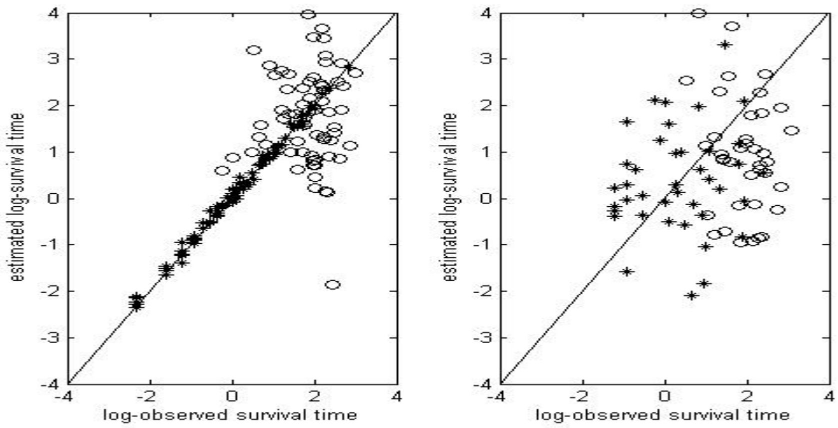

Figure 5.1 Scatter plots of estimated log-survival times versus log-observed survival times

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, **19**, 716-723.

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gGene expression data. *PLoS Biology*, **2**, 511-522.

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

Cho, D. H., Shim, J. and Seok, K. H. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of the Korean Data & Information Science Society*, **21**, 155-162.

Cox, D. R. (1972) Regression models and life tables (with discussions). *Journal of the Royal Statistical Society B*, **74**, 187-220.

Gehan, E. A. (1965). Generalized Wilcoxon test for comparing arbitrarily singe-censored samples. *Biometrika*, **52**, 202-223.

Ghosh, K. S. and Ghosal, S. (2006). Semiparametric accelerated failure time models for censored data. *Bayesian Statistics and Its Applications*, **15**, 213-229.

Hu, S. and Rao, J. S. (2010). *Sparse penalization with censoring constraints for estimating high dimensional AFT models with applications to microarray data analysis,* Technical Report 07 of Division of Biostatistics, Case Western Reserve University, OH.

Huang, J., Ma, S. and Xie, H. (2005). *Regularized estimation in the accelerated failure time model with high dimensional covariates,* Technical Report No. 349, Department of Statistics and Actuarial Science, The University of Iowa, IA.

Hwang, C. and Shim, J. (2010). Semiparametric support vector machine for accelerated failure time model. *Journal of the Korean Data & Information Science Society*, **21**, 467-477.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, **90**, 341-353.

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

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

Krishnapuram, B., Carlin, L., Figueiredo, M. A. T. and Hartermink, A. J. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 957-968.

Li, H. (2006). *Censored data regression in high-dimension and low-sample size settings for genomic applications,* UPenn Biostatistics Working Paper 9, University of Pennsylvania, PA.

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

Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B., Giltnane J. M. and et al. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*, **346**, 1937-1947.

Sauerbrei, W. and Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistical Medicine*, **11**, 2093-2099.

Shim, J. (2005). Censored kernel ridge regression. *Journal of the Korean Data & Information Science Society*, **16**, 1045-1052.

Shim, J. and Lee, J. T. (2009). Kernel method for autoregressive data. *Journal of the Korean Data & Information Science Society*, **20**, 467-472.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, **58**, 267-288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, **16**, 385-395.

Williams, P. M. (1995). Bayesian regularization and pruning using a Laplace prior. *Neural Computation*, **7**, 117-143.

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

Zhou, M. (1998). *Regression with censored data : The synthetic data and least squares approach,* Technical Report 374, University of Kentucky, KY.