Sparse Estimation and Inference for Censored Median Regression

Justin Hall Shows, **Wenbin Lu**, and **Hao Helen Zhang**
Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

Abstract

Censored median regression has proved useful for analyzing survival data in complicated situations, say, when the variance is heteroscedastic or the data contain outliers. In this paper, we study the sparse estimation for censored median regression models, which is an important problem for high dimensional survival data analysis. In particular, a new procedure is proposed to minimize an inverse-censoring-probability weighted least absolute deviation loss subject to the adaptive LASSO penalty and result in a sparse and robust median estimator. We show that, with a proper choice of the tuning parameter, the procedure can identify the underlying sparse model consistently and has desired large-sample properties including root-*n* consistency and the asymptotic normality. The procedure also enjoys great advantages in computation, since its entire solution path can be obtained efficiently. Furthermore, we propose a resampling method to estimate the variance of the estimator. The performance of the procedure is illustrated by extensive simulations and two real data applications including one microarray gene expression survival data.

1. Introduction

In the last forty years, a variety of semiparametric survival models have been proposed for censored time-to-event data analysis. Among them, the proportional hazards model (Cox, 1972) is perhaps the most popular one and widely used in literature. However, as noted by many authors, the proportional hazards assumption may not be appropriate in some practical applications; for some problems alternative models may be more suitable. One of the attractive alternatives is the accelerated failure time (AFT) model (Kalbfleisch and Prentice, 1980; Cox and Oakes, 1984), which relates the logarithm of the failure time linearly to covariates as

$$\log(T) = \beta'Z + \varepsilon, \tag{1}$$

where T is the failure time of our interest, Z is the p-dimensional vector of covariates and ε is the error term with a completely unspecified distribution that is independent of Z. Due to the linear model structure, the results of the AFT model are easier to interpret (Reid, 1994). Due to these desired properties, the AFT model has been extensively studied in the literature (Prentice, 1978; Buckley and James, 1979; Ritov, 1990; Tsiatis, 1990; Ying, 1993; Jin et al., 2003; among others).

A major assumption of the AFT model is that the error terms are i.i.d., which is often too restrictive in practice (Koenker and Geling, 2001). To handle more complicated problems, say, data with heteroscedastic or heavy-tailed error distributions, censored quantile regression

provides a natural remedy. In the context of survival data analysis, Ying et al. (1995), Bang and Tsiatis (2003), and Zhou (2006), among others, considered median regression with random censoring and derived various inverse-censoring-probability weighted methods for parameter estimation; Portnoy (2003) considered a general censored quantile regression by assuming that all conditional quantiles are linear functions of the covariates and developed recursively re-weighted estimators of regression parameters. In general, quantile regression is more robust against outliers and may be more effective for nonstandard data; the regression quantiles may be well estimated even if the mean is not for censored survival data.

In many biomedical studies, a large number of predictors are often collected but not all of them are relevant for the prediction of survival times. In other words, the underlying true model has a sparse model structure in terms of the appearance of covariates in the model. For example, many cancer studies start measuring patients' genomic information and combine it with traditional risk factors to improve cancer diagnosis and make personalized drugs. Genomic information, often in the form of gene expression patterns, is high dimensional and typically only a small number of genes contains relevant information, so the underlying model is naturally sparse. How to identify this small group of relevant genes is an important yet challenging question. Simultaneous variable selection and parameter estimation then becomes crucial since it helps to produce a parsimonious model with better risk assessment and model interpretation. There is a rich literature for variable selection in standard linear regression and survival data analysis based on the proportional hazards model. Traditional procedures include backward deletion, forward addition and stepwise selection. However, these procedures may suffer from high variability (Breiman, 1996). Recently some shrinkage methods have been proposed based on the penalized likelihood or partial likelihood estimation, including the LASSO (Tibshirani, 1996, 1997), the SCAD (Fan and Li, 2001, 2002) and the adaptive LASSO (Zou 2006, 2008; Zhang and Lu, 2007; Lu and Zhang, 2007). More recently, several authors extend the shrinkage methods to variable selection of the AFT model using the penalized rank estimation or Buckley-James's estimation (Johnson, 2008; Johnson et al., 2008; Cai et al., 2008).

To the best of our knowledge, the sparse estimation of censored quantile regression has not been studied in the literature. When there is no censoring, Wang et al. (2007) considered the LAD-LASSO estimation for linear median regression assuming i.i.d. error terms. For nonparametric and semiparametric regression quantile models, the total variation penalty has been proposed and studied extensively, including Koenker et al. (1994), He and Shi (1996), He, Ng and Portnoy, S. (1998), He and Ng (1999), and some recent work on penalized triograms by Koenker and Mizera (2004) and on longitudinal models by Koenker (2004). The L_1 nature of the total variation penalty make it work like a model selection process for identifying a small number of spline basis functions. In this paper, we focus on the sparse estimation of censored median regression. Compared with variable selection for standard linear regression models, there are two additional challenges in censored median regression: the error distributions are not i.i.d. and the response is censored. To tackle these issues, we construct the new estimator based on an inverse-censoring-probability weighted least absolute deviation estimation method subject to some shrinkage penalty. To facilitate the selection of variables, we suggest the use of the adaptive LASSO penalty, which is shown to yield an estimator with desired properties both in theory and in computation. The proposed method can be easily extended to arbitrary quantile regression with random censoring. The remainder of the article is organized as follows. In Section 2, we develop the sparse censored median regression (SCMR) estimator for right-censored data and establish its large-sample properties, such as root-n consistency, sparsity, and asymptotic normality. Section 3 discusses the computational algorithm and solution path, and the tuning parameter issue is addressed as well. We also derive a bootstrap method for estimating the variance of the SCMR estimator. Sections 4 is devoted to simulations and real

examples. Final remarks are given in Section 5. Major technical derivations are contained in the Appendix.

2. Sparse censored median regression

2.1 The SCMR estimator

Consider a study of n subjects. Let $\{T_i, C_i, Z_i, i = 1, \ldots, n\}$ denote n triplets of failure times, censoring times, and p-dimensional covariates of interest. Conditional on Z_i, the median regression model assumes

$$\log(T_i) = \theta_0' X_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2}$$

where $X_i = \left(1, Z_i'\right)'$, $\theta_0 = \left(\alpha_0, \beta_0'\right)'$ is a $(p+1)$-dimensional vector of regression parameters and ε_i's are assumed to have a conditional median of 0 given covariates. Here we do not require the i.i.d. assumption for ε_i's as in Wang et al. (2007). Note that the log transformation in (2) can be replace by any specified monotonic transformation. Define $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, then the observed data consist of $(\tilde{T}_i, \delta_i, Z_i, i = 1, \ldots, n)$. As in Ying et al. (1995) and Zhou (2006), we assume that the censoring time C is independent of T and Z. Let $G(\cdot)$ denote the survival function of censoring times. To estimate the parameters θ_0, Ying et al. (1995) proposed to solve

$$\sum_{i=1}^{n} X_i \left[\frac{I\{\log(\tilde{T}_i) - \theta' X_i \geq 0\}}{\widehat{G}(\theta' X_i)} - \frac{1}{2} \right] = 0,$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator of $G(\cdot)$ based on the data $(\tilde{T}_i, 1 - \delta_i)$, $i = 1, \ldots, n$. Note that the above equation is neither continuous nor monotone in θ and thus is difficult to solve especially when the dimension of θ is high. Instead, Bang and Tsiatis (2003) proposed an alternative inverse censoring-probability weighted estimating equation as follows:

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} X_i \left[I\{\log(\tilde{T}_i) - \theta' X_i \geq 0\} - \frac{1}{2} \right] = 0.$$

As pointed out by Zhou (2006), the solution of the above equation is also a minimizer of

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| \log(\tilde{T}_i) - \theta' X_i \right|, \tag{3}$$

which is convex in θ and can be easily solved using an efficient linear programming algorithm of Koenker and D'Orey (1987). In addition, Zhou (2006) showed that the above inverse censoring probability weighted least absolute deviation estimator has better finite sample performance than that of Ying et al. (1995).

To conduct variable selection for censored median regression, we propose to minimize

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| \log(\tilde{T}_i) - \theta' X_i \right| + n \sum_{j=1}^{p} J_\lambda(|\beta_j|), \tag{4}$$

where the penalty J_λ is chosen to achieve simultaneous selection and coefficient estimation. Two popular choices include the LASSO penalty $J_\lambda(|\beta_j|) = \lambda|\beta_j|$ and the SCAD penalty (Fan and Li, 2001). In this paper, we consider an alternative regularization framework using the adaptive LASSO penalty (Zou, 2006; Zhang and Lu, 2007; Wang et al., 2007; among others). Specifically, let $\tilde{\theta} \equiv (\tilde{\alpha}, \tilde{\beta}')'$ denote the minimizer of (3). Our SCMR estimator is defined as the minimizer of

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| \log(\tilde{T}_i) - \theta'X_i \right| + n\lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j|},$$

(5)

where $\lambda > 0$ is the tuning parameter and $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)'$. Note the estimators from (3) are used as the weights to leverage the LASSO penalty in (5). The the adaptive LASSO procedure has been shown to achieve consistent variable selection asymptotically in various scenarios when the tuning parameter λ is chosen properly.

2.2 Theoretical properties of SCMR estimator

Suppose that $\beta_0 = \left(\beta_{a,0}', \beta_{b,0}'\right)'$ and is arranged so that $\beta_{a,0} = (\beta_{10}, \dots, \beta_{q0})' \neq 0$ and $\beta_{b,0} = (\beta_{(q+1)0}, \dots, \beta_{p0})' = 0$. Here it is assumed that the length of $\beta_{a,0}$ is q. Correspondingly, denote the minimizer of (5) by $\widehat{\theta} = \left(\widehat{\alpha}, \widehat{\beta}_a', \widehat{\beta}_b'\right)'$. Define

$\theta_{a,0} = \left(\alpha_0, \beta_{a,0}'\right)', \widehat{\theta}_a = \left(\widehat{\alpha}, \widehat{\beta}_a'\right)'$ and $b_1 = \left(\dfrac{sign(\beta_1)}{|\beta_1|}, \dots, \dfrac{sign(\beta_q)}{|\beta_q|}\right)'$. Under the regularity conditions given in the Appendix, we have the following two theorems.

Theorem 1—$\left(\sqrt{n}\text{–Consistency}\right)$ *If* $\sqrt{n}\lambda = O_p(1)$, *then* $\sqrt{n}\|\widehat{\theta} - \theta_0\| = O_p(1)$.

Theorem 2—*If* $\sqrt{n}\lambda \to \lambda_0 \geq 0$ *and* $n\lambda \to \infty$, *then as* $n \to \infty$

i. *(Selection-Consistency)* $P(\hat{\beta}_b = 0) \to 1$.

ii. *(Asymptotic Normality)* $\sqrt{n}\left(\widehat{\theta}_a - \theta_{a,0}\right)$ *converges in distribution to a normal random vector with mean* $-\Sigma_a'\lambda_0 b_1$ *and variance-covariance matrix* $\Sigma_a^{-1} V_a \Sigma_a^{-1}$.

The definitions of Σ_a and V_a, and the proofs of Theorems 1 and 2 are given in the Appendix.

3. Computation, parameter tuning and variance estimation

3.1. Computational algorithm using linear programming

Given a fixed tuning parameter λ, we suggest the following algorithm to compute the SCMR estimator. First, we reformulate (5) as a least absolute deviation problem without penalty. To be specific, we define,

$$(X^*)' = \begin{pmatrix} \frac{\delta_1}{\widehat{G}(\tilde{T}_1)} & \frac{\delta_1}{\widehat{G}(\tilde{T}_1)} X_{11} & \cdots & \frac{\delta_1}{\widehat{G}(\tilde{T}_1)} X_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\delta_n}{\widehat{G}(\tilde{T}_n)} & \frac{\delta_n}{\widehat{G}(\tilde{T}_n)} X_{n1} & \cdots & \frac{\delta_n}{\widehat{G}(\tilde{T}_n)} X_{np} \\ 0 & \frac{n\lambda}{|\tilde{\beta}_1|} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \frac{n\lambda}{|\tilde{\beta}_p|} \end{pmatrix}_{(n+p)\times(p+1)} ,$$

$$Y^* = \left(\frac{\delta_1}{\widehat{G}(\tilde{T}_1)} \log(\tilde{T}_1), \cdots, \frac{\delta}{\widehat{G}(\tilde{T}_n)} \log(\tilde{T}_n), 0, \cdots, 0 \right)'_{(n+p)\times 1} .$$

Then, with λ fixed, minimizing (5) is equivalent to minimizing

$$\sum_{i=1}^{n} \left| Y^* - \theta' X_i^* \right|,$$

which can be easily computed using any statistical software for linear programming, for example, the *rq* function in R.

3.2. Solution path

Efron et al. (2004) proposed an efficient path algorithm called LARS to get the entire solution path of LASSO, which can greatly reduce the computation cost of model fitting and tuning. Recently, Li and Zhu (2008) modified the LARS algorithm to get the solution paths of the sparse quantile regression with uncensored data. It turns out that the SCMR estimator also has a solution path linear in λ. In order to compute the entire solution path of SCMR, we first modify the algorithm of Li and Zhu (2008) to compute the solution path of a weighted quantile regression subject to the LASSO penalty

$$\sum_{i=1}^{n} w_i \left| Y_i - \theta' X_i \right| + n\lambda \sum_{j=1}^{p} \left| \beta_j \right| .$$

(6)

In our SCMR estimators, the weight $w_i = \delta_i / \hat{G}(\tilde{T}_i)$. The corresponding R codes are available from the authors.

Next we propose the following algorithm to compute the solution path of the SCMR estimator with the adaptive LASSO penalty.

Step 1: Create working data $X_i^{**} \equiv \left(1, Z_{i1}^*, \ldots, Z_{ip}^* \right)'$ with $Z_{ij}^* = Z_{ij} |\tilde{\beta}_j|$ for $i = 1, \ldots, n; j = 1, \ldots, p$.

Step 2: Applying the solution path algorithm for (6), solve

$$\hat{\gamma} = \arg \min_\gamma \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| Y - \gamma' X_i^{**} \right| + n\lambda \sum_{j=2}^{p+1} \left| \gamma_j \right| .$$

Step 3: Define $\hat{\alpha} = \hat{\gamma}_1$ and $\hat{\beta}_j = \hat{\gamma}_{j+1}|\tilde{\beta}_j|$ for $j = 1, \ldots, p$

.

Note that the above two computation methods give exactly the same solutions for a common λ. But the solution path method is more effective since it produce the whole solution path, while the first method is more convenient to combine with the bootstrap method to compute the variance of the SCMR estimator for a given λ.

3.3. Parameter tuning

Our estimator makes use of the adaptive LASSO penalty, i.e., a penalty of the form

$\lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j|}$, where $\tilde{\beta}$ is the unpenalized LAD estimator. We propose to tune the parameter λ based on a BIC-type criteria. Note that if there is no censoring and the error terms are i.i.d., the least absolute deviation loss is closely related to linear regression with double exponential error. To derive the BIC-type tuning procedure, we make use of this connection. More specifically, assume that the error term ε_i's are i.i.d. double exponential variables with the scale parameter σ. When there is no censoring, up to a constant the negative log likelihood function can be written as $\sum_{i=1}^{n} |\log(T_i) - \theta'X_i|/\sigma$ and the maximum likelihood estimator of σ is given by $(1/n) \sum_{i=1}^{n} |\log(T_i) - \tilde{\theta}'X_i|$ with $\tilde{\theta}$ being the LAD estimator. Motivated by this observation, we propose the following tuning procedure for sparse censored median regression:

$$BIC(\lambda) = \frac{2}{\tilde{\sigma}} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| \log(\tilde{T}_i) - X_i' \widehat{\theta}(\lambda) \right| + \log(n)r,$$

where $\tilde{\sigma} = (1/n) \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \left| \log(\tilde{T}_i) - X_i' \tilde{\theta} \right|$ and r is the number of non-zero elements in $\hat{\beta}(\lambda)$. Then we minimize the above criteria over a range of values of λ for choosing the best tuning parameter.

3.4. Variance estimation

Next, we propose a bootstrap method to estimate the variance of our estimator. Specifically, we first take a large number of random samples of size n (with replacement) from the observed data. Then for the kth bootstrapped data $\left(\tilde{T}_i^{(k)}, \delta_i^{(k)}, X_i^{(k)}, i=1,\ldots,n\right)$, we compute the Kaplan-Meier estimator $\hat{G}^{(k)}(\cdot)$ and the inverse censoring probability weighted least absolute deviation estimator $\tilde{\theta}^{(k)}$. Lastly, we compute the bootstrapped estimate $\hat{\theta}^{(k)}$ by minimizing

$$\sum_{i=1}^{n} \frac{\delta_i^{(k)}}{\widehat{G}^{(k)}(\tilde{T}_i^{(k)})} \left| \log(\tilde{T}_i^{(k)}) - \theta'X_i^{(k)} \right| + n\lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j^{(k)}|}.$$

Here for saving the computation cost, we fixed λ at the optimal value chosen by each tuning method based on the original data.

4. Numerical studies

4.1 Simulation studies

In this section, we examine the finite sample performance of the proposed SCMR estimator in terms of variable selection and model estimation. In addition, we conduct a series of sensitivity analyses to check the performance of our estimator when the random censoring assumption is violated.

The failure times are generated from the median regression model (2). Four error distributions are considered: $t(5)$ distribution, double exponential distribution with median 0 and scale parameter 1, standard extreme value distribution, and standard logistic distribution. Since the standard extreme value distribution has median $\log(\log(2))$ rather than 0, we subtract its median. We consider eight covariates, which are i.i.d. standard normal random variables. The regression parameter vector is chosen as $\beta_0 = (0.5, 1, 1.5, 2, 0, 0, 0, 0)$ with the intercept $\alpha_0 = 1$. The censoring times are generated from a uniform distribution on $(0, c)$, where c is a constant to obtain the desired censoring level. We consider two censoring rates: 15% and 30%. For each censoring rate, we consider samples of sizes 50, 100, and 200. For each setting, we conduct 100 runs of simulations.

For comparisons, we fit six different estimators in our analysis, namely the full model estimator $\tilde{\beta}$ (called "Full" in the Tables), our SCMR estimators with the adaptive Lasso penalty ("SCMR-A"), the Lasso penalty ("SCMR-L"), and the SCAD penalty ("SCMR-S"), the naive estimator with the adaptive Lasso penalty ("Naive") in which the censoring information was ignored and all the observed event times were treated as failure times, and the oracle estimator assuming the true model were known ("Oracle"). To obtain the SCMR-S estimator, we use the local linear approximation (LLA) algorithm by Zou and Li (2008). Various estimators are compared with regard to their overall mean absolute deviation (MAD), point estimation accuracy, and the variable selection performance. Here the MAD of an estimator $\hat{\theta}$ is given by

$$MAD = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{\theta}' X_i - \theta_0' X_i \right|.$$

We summarize the overall model estimation and variable selection results of different estimators in Tables 1 and 3, in terms of the MAD, the selection frequency of each variable, the frequency of selecting the exact true model, and the mean number of incorrect and correct zeros selected over 100 simulation runs. To simplify the presentation, we only present the results for $t(5)$ and double exponential error distributions; the results for other error distributions have similar patterns and hence omitted. In terms of model estimation, the MAD of the SCMR-A and SCMR-S estimators are smaller than those of the SCMR-L and Full estimators in all the settings, and are also close to the oracle estimator when the sample size is large and the censoring proportion is low. With regard to variable selection, compared with the Full and SCMR-L estimators, the SCMR-A and SCMR-S estimators produce more sparse models and select the exact true model more frequently. For example, when $n = 200$, censoring rate is 15% and the error distribution is $t(5)$, the frequencies of selecting true model (SF) are: Oracle (100), SCMR-A (72), SCMR-L (22), SCMR-S (72), Naive (69), and Full (0). In average, the numbers of correct/incorrect zeros in the estimators are: Oracle (4/0), SCMR-A (3.67/0), SCMR-L (2.59/0), SCMR-S (3.63/0), Naive (3.65/0), and Full (0/0). In summary, the SCMR-A and SCMR-S estimators give comparable performance in both estimation and variable selection, which is not surprising since they are both selection consistenct and have the same asymptotic distribution. In addition, the naive estimator has the biggest MAD in all settings since it is a biased estimator, but it is interesting to note that it selects variables well.

The point estimation results of the first three non-zero coefficients are summarized in Tables 2 and 4, including the bias and sample standard deviation of the estimates over 100 runs, and the mean of the estimated standard errors based on 500 bootstraps per run. Based on the results, the naive estimator showed big biases as we expected; the biases of all other estimators are relatively small particularly when the sample size is large and the censoring proportion is low. In addition, the estimated standard errors obtained using the proposed bootstrap method are reasonably close to the sample standard deviations in all scenarios.

Next, we conduct a series of sensitivity analyses to check the performance of our estimator when the random censoring assumption is violated. More specifically, the censoring times are now generated as $C_i = e^{\xi X_{1i}} C_i^*$, $i = 1, \dots, n$, where C_i^*'s are from a uniform distribution on $(0, c)$ as before. We consider two values of $\xi : \xi = 0.1$ or 0.2, two censoring rates: 15% and 50%, and two sample sizes: $n = 100$ or 500. Other settings remain the same as before. The estimation and variable selection results are for $\xi = 0.2$ with $t(5)$ error distribution are summarized in Tables 5 and 6, respectively. The results for $\xi = 0.1$ and other error distributions are quite similar, which are omitted here. Based on the results reported in Tables 5 and 6, we observe similar findings as before. The SCMR-A estimator showed much better performance in terms of variable selection compared with the SCMR-L and the Full estimators, although all the methods may produce certain biases in point estimation as expected.

4.2 PBC data

The primary biliary cirrhosis (PBC) data was collected at the Mayo clinic between 1974 and 1984. This data is given in Therneau and Grambsch (2000). In this study, 312 patients from a total of 424 patients who agreed to participate in the randomized trial are eligible for the analysis. Of those, 125 patients died before the end of follow-up. We study the dependence of the survival time on the following selected covariates: (1) continuous variables: age (in years), alb (albumin in g/dl), alk (alkaline phosphatase in U/liter), bil (serum bilirunbin in mg/dl), chol (serum cholesterol in mg/dl), cop (urine copper in µg/day), plat (platelets per cubic ml/1000), prot (prothrombin time in seconds), sgot (liver enzyme in U/ml), trig (triglycerides in mg/dl); (2) categorical variables: asc (0, absence of ascites; 1, presence of ascites), ede (0 no edema; 0.5 untreated or successfully treated; 1 unsuccessfully treated edema), hep (0, absence of hepatomegaly; 1, presence of hepatomegaly), sex (0 male; 1 female), spid (0, absence of spiders; 1, presence of spiders), stage (histological stage of disease, graded 1, 2, 3 or 4), trt (1 control, 2 treatment).

The PBC data has been previously analyzed by a number of authors using various estimation and variable selection methods. For example, Tibshirani (1997) fitted the proportional hazards model with the stepwise selection and with the LASSO penalty. Zhang and Lu (2007) further studied the PBC data using the penalized partial likelihood estimation method with the SCAD and the adaptive Lasso penalty. Here, we apply the proposed SCMR method to the PBC data. As in Tibshirani (1997) and Zhang and Lu (2007), we restrict our attention to the 276 observations without missing values. Among these 276 patients, there are 111 deaths, about 60% of censoring. Table 7 summarizes the estimated coefficients and the standard errors for the Full, the SCMR-L, the SCMR-A and the SCMR-S. We found that the SCMR-A selects 9 variables: *age, asc, oed, bil, alb, cop, alk, plat* and *prot* and the SCMR-L selects 13 variables which contain the 9 variables selected by the SCMR-A. Moreover, the 9 variables selected by SCMR-A shared 6 variables out of 8 selected by the penalized partial likelihood estimation method of Zhang and Lu (2007) in the proportional hazards model. We also plot the solution path of the SCMR-A estimator in Figure 1.

4.3 DLBCL microarray data

Sparse model estimation has wide applications in high dimensional data analysis. In this example, we apply the SCMR method to the high dimensional microarray gene expression data of Rosenwald et al. (2002). The data consists of survival times of 240 diffuse large B-cell lymphoma (DLBCL) patients, and the expressions of 7, 399 genes for each patient. Among them, 138 patients died during the follow-up method. The main goals of the study are to identify the important genes that can predict patients' survival and to study their effects on survival. This data was analyzed by Li and Luan (2005). To handle such high dimensional data, a common practice is to first conduct a preliminary gene filtering based on some univariate analysis, and then apply a more sophisticated model-based analysis. Following Li and Luan (2005), we concentrate on the top 50 genes selected using the univariate log-rank test. To evaluate the performance of the proposed SCMR method, the data are randomly divided into two sets: the training set (160 patients) and the testing set (80 patients).

The SCMR-A estimator is then computed based on the training data and the proposed BIC method is used for parameter tuning. Our SCMR-A estimator selects totally 25 genes. To evaluate the prediction performance of the resulting SCMR-A estimator built with the training set, we plot, in Figure 2, the Kaplan-Meier estimates of survival functions for the high-risk and low-risk groups of patients, defined by the estimated conditional medians of failure times. The cut-off value was determined by the median failure time of the baseline group from the training set, and the same cutoff was applied to the testing data. It is seen that the separation of the two-risk groups is reasonably good in both the training and the testing data, suggesting a satisfactory prediction performance of the fitted survival model. The log-rank test of differences between two survival curves gives p-values of 0 and 0.031 for the training and testing data, respectively.

5. Discussion

We propose the SCMR estimator for model estimation and selection in censored median regression based on the inverse-censoring-probability-weighted method combing with the adaptive LASSO penalty. Theoretical properties, such as variable selection consistency and asymptotic normality, of the SCMR estimator are established. The whole solution path of the SCMR estimator can be obtained and the proposed method for censored median regression can be easily generalized to censored quantile regression.

A key assumption of the proposed method is that the censoring distribution is independent of covariates. However, such a restrictive assumption can be relaxed. For example, Wang and Wang (2009) recently proposed a locally weighted censored quantile regression, in which a local Kaplan-Meier estimator was used to estimate the conditional survival function of censoring times given covariates. We think that the proposed SCMR estimator can be easily generalized to accommodate this case. However, when the dimension of covariates is high, the local Kaplan-Meier estimator may not work due to the curse of dimensionality. Alternatively, Portnoy (2003) and Peng and Huang (2008) studied a class of censored quantile regressions at all quantile levels under the weaker assumption that censoring and survival times are conditionally independent. The model parameters are estimated through a series of estimating equations: the self-consistency equations in Portnoy (2003) and the martingale-based equations in Peng and Huang (2008). The variable selection for such estimating equation based methods becomes more challenging, which needs further investigation.

Appendix: Proof of Theorems

To prove the asymptotic results established in Theorems 1 and 2, we need the following regularity conditions:

1. The error term ε has a continuous conditional density $f(\cdot|Z = z)$ satisfying that $f(0|Z = z) \geq b0 > 0$, $|\dot{f}(0|Z = z)| \leq B_0$ and $\sup_s f(s|Z = z) \leq B_0$ for all possible values z of Z, where (b_0, B_0) are two positive constants and $\dot{f}$ is the derivative of f.

2. The covariate vector Z are of compact support and the parameter β_0 belongs to the interior of a known compact set $\mathcal{B}_0$.

3. $P(t \leq T \leq C) \geq \zeta_0 > 0$ for any $t \in [0, \tau]$, where τ is the maximum follow-up and ζ_0 is a positive constant.

PROOF OF THEOREM 1

To establish the result given in Theorem 1, it is equivalent to show that for any $\eta > 0$, there is a constant M such that $P\left(\sqrt{n}\|\widehat{\theta} - \theta\| \leq M\right) \geq 1 - \eta$. Let $u = (u_0, u_1, \ldots, u_p)' \in \Re^{p+1}$, and $A_M = \{\theta_0 + \frac{u}{\sqrt{n}} : \|u\| \leq M\}$ be a ball in $\Re^{p+1}$ centered at θ_0 with the radius $M/\sqrt{n}$. Then we need to show that $P(\widehat{\theta} \in A_M) \geq 1 - \eta$. Define

$$Q(G, \theta) = \sum_{i=1}^{n} \frac{\delta_i}{G(\tilde{T}_i)} \left|\log(\tilde{T}_i) - \theta' X_i\right| + n\lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\tilde{\beta}_j|},$$

which is a convex function of θ. Thus, to prove $P(\widehat{\theta} \in A_M) \geq 1 - \eta$ for any $\eta > 0$, it is sufficient to show

$$P\left\{\inf_{\|u\| = M} Q\left(\widehat{G}, \theta_0 + \frac{u}{\sqrt{n}}\right) > Q\left(\widehat{G}, \theta_0\right)\right\} \geq 1 - \eta.$$

Let $D_n(u) = Q\left(\widehat{G}, \theta_0 + \frac{u}{\sqrt{n}}\right) - Q\left(\widehat{G}, \theta_0\right)$, which can be written as

$$\left\{Q(G_0, \theta_0 + \frac{u}{\sqrt{n}}) - Q(G_0, \theta_0)\right\} + \left\{Q(\widehat{G}, \theta_0 + \frac{u}{\sqrt{n}}) - Q(G_0, \theta_0 + \frac{u}{\sqrt{n}})\right\} - \left\{Q(\widehat{G}, \theta_0) - Q(G_0, \theta_0)\right\}. \tag{7}$$

where $G_0(\cdot)$ is the true survival function of the censoring time.

For the first term in (7), we have

$$Q(G_0, \theta_0 + \tfrac{u}{\sqrt{n}}) - Q(G_0, \theta_0)$$
$$= \sum_{i=1}^{n} \frac{\delta_i}{G_0(\tilde{T}_i)} \left(|\varepsilon_i - X_i'\frac{u}{\sqrt{n}}| - |\varepsilon_i|\right) + n\lambda \sum_{j=1}^{p} \frac{|\beta_{j0} + u_j/\sqrt{n}| - |\beta_{j0}|}{|\tilde{\beta}_j|}$$
$$\geq \sum_{i=1}^{n} \frac{\delta_i}{G_0(\tilde{T}_i)} \left(|\varepsilon_i - X_i'\frac{u}{\sqrt{n}}| - |\varepsilon_i|\right) + n\lambda \sum_{j=1}^{q} \frac{|\beta_{j0} + u_j/\sqrt{n}| - |\beta_{j0}|}{|\tilde{\beta}_j|}$$
$$\geq \sum_{i=1}^{n} \frac{\delta_i}{G_0(\tilde{T}_i)} \left(\left|\varepsilon_i - X_i'\frac{u}{\sqrt{n}}\right| - |\varepsilon_i|\right) - \sqrt{n}\lambda \sum_{j=1}^{q} \frac{|u_j|}{|\tilde{\beta}_j|}$$
$$\equiv L_n(u) - \sqrt{n}\lambda \sum_{j=1}^{q} \frac{|u_j|}{|\tilde{\beta}_j|} \geq L_n(u) - \kappa_1 Q_p(\|u\|).$$

where κ_1 is a finite positive constant. The last inequality in the above expression is because that $\sqrt{n}\lambda = O_p(1)$ and $\tilde{\beta}_j, j = 1, \ldots, q$, converges to β_{j0} that is bounded away from zero.

In addition, based on the result (Knight, 1998) that for any

$$x \neq 0, |x - y| - |x| = -y\,[I(x>0) - I(x<0)] + 2\int_0^y [I(x \leq s) - I(x \leq 0)]\,ds,\text{ we have}$$

$$L_n(u) = \frac{u'}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{G_0(\tilde{T}_i)} X_i \{I(\varepsilon_i<0) - I(\varepsilon_i>0)\} + 2\sum_{i=1}^n \frac{\delta_i}{G_0(\tilde{T}_i)} \int_0^{u'X_i/\sqrt{n}} [I(\varepsilon_i \leq s) - I(\varepsilon_i \leq 0)]\,ds.$$

Since ε_i has median zero, it is easy to show that

$$E\left[\frac{\delta_i}{G_0(\tilde{T}_i)} X_i \{I(\varepsilon_i<0) - I(\varepsilon_i>0)\}\right] = 0.$$

Thus, by the central limit theorem, $\dfrac{u'}{\sqrt{n}} \sum_{i=1}^n \dfrac{\delta_i}{G_0(\tilde{T}_i)} X_i \{I(\varepsilon_i<0) - I(\varepsilon_i>0)\}$ converges in distribution to $u'W_1$, where W_1 is a $(p + 1)$-dimensional normal with mean 0 and variance-covariance matrix $\Sigma_1 = E\left(\dfrac{X_1 X_1'}{G_0(\tilde{T}_1)}\right)$. It implies that the first term in $L_n(u)$ can be written as $O_p(\|u\|)$. For the second term in $L_n(u)$, let

$$A_{ni}(u) = \frac{\delta_i}{G_0(\tilde{T}_i)} \int_0^{u^T X_i/\sqrt{n}} [I(\varepsilon_i \leq s) - I(\varepsilon_i \leq 0)]\,ds.$$ We will show that $\sum_{i=1}^n A_{ni}(u)$ converges in probability to a quadratic function of u. More specifically, for any $\psi > 0$, write

$$A_{ni}^2(u) = A_{ni}^2(u) I\left(\frac{|u'X_i|}{\sqrt{n}} \geq \psi\right) + A_{ni}^2(u) I\left(\frac{|u'X_i|}{\sqrt{n}} < \psi\right).$$

We have

$$nE\{A_{ni}^2(u) I(|u'X_i| \geq \psi\sqrt{n})\}$$
$$\leq nE\left\{\frac{\delta_i}{G_0^2(\tilde{T}_i)} \left(\int_0^{u'X_i/\sqrt{n}} 2ds\right)^2 I(|u^T X_i| \geq \psi\sqrt{n})\right\}$$
$$= \frac{4}{s_0} E\{|u'X_i|^2 I(|u'X_i| \geq \psi\sqrt{n})\} \to 0,\text{ as } n \to \infty.$$

Moreover,

$$nE\{A_{ni}^2(u) I(|u'X_i| < \psi\sqrt{n})\}$$
$$\leq nE\left(\frac{\delta_i}{G_0^2(\tilde{T}_i)} 2\int_0^{|u'X_i|/\sqrt{n}} ds \int_0^{|u'X_i|/\sqrt{n}} \{I(\varepsilon_i \leq s) - I(\varepsilon_i \leq 0)\} ds I(|u'X_i| < \psi\sqrt{n})\right)$$
$$\leq \frac{2n\psi}{s_0} E_z\left[\int_0^{|u'X_i|/\sqrt{n}} \{F(s|Z) - F(0|Z)\} ds I(|u'X_i| < \psi\sqrt{n})\right]$$
$$\leq \frac{2n\psi}{s_0} E_z\left\{\int_0^{|u'X_i|/\sqrt{n}} f(s^*|Z) s\, ds I(|u'X_i| < \psi\sqrt{n})\right\}$$
$$\leq \frac{\psi B_0}{s_0} E_z\{|u'X_i|^2 I(|u'X_i| < \psi\sqrt{n})\} \leq \frac{\psi B_0}{s_0} E_z(|u'X_i|^2).$$

Since $E_Z(|u'X_i|^2)$ is bounded and ψ can be arbitrary small, it follows that

$\frac{\psi B_0}{\zeta_0} E_z\left(|u'X_i|^2\right) \to 0$ as $\psi \to 0$. Thus, we have, as $n \to \infty$,

$$Var\{\sum_{i=1}^n A_{ni}(u)\}=\sum_{i=1}^n Var\{A_{ni}(u)\} \le nE\{A_{ni}^2(u)\} \to 0,$$

which implies $\sum_{i=1}^n [A_{ni}(u) - E\{A_{ni}(u)\}] = o_p(1)$. Furthermore, we have

$$\sum_{i=1}^n E\{A_{ni}(u)\} \quad = nE\left[\frac{\delta_i}{G_0(\tilde{T}_i)}\int_0^{u'X_1/\sqrt{n}}\{I(\varepsilon_1 \le s) - I(\varepsilon_1 \le 0)\}ds\right]$$

$$=nE_z\left[\int_0^{u'X_i/\sqrt{n}}\{F(s|Z) - F(0|Z)\}ds\right]$$

$$=nE_z\left\{\int_0^{u'X_i/\sqrt{n}}sf(0|Z)ds\right\}+o_p(1)$$

$$=\frac{1}{2}u'\Sigma u+o_p(1),$$

where $\Sigma = E_Z\{f(0|X)XX'\}$ is positive and finite. Thus,

$$Q(G_0,\theta_0+\frac{u}{\sqrt{n}}) - Q(G_0,\theta_0) \ge \frac{1}{2}u'\Sigma u+u'W_1 - \kappa_1 O_p(\|u\|)+o_p(1).$$

For the second and third terms in (7), by the Taylor expansion, we have

$$\sqrt{n}\{\frac{1}{\widehat{G}(\tilde{T}_i)} - \frac{1}{G_0(\tilde{T}_i)}\} \quad = -\frac{\sqrt{n}\{\widehat{G}(\tilde{T}_i)-G_0(\tilde{T}_i)\}}{G_0^2(\tilde{T}_i)}+o_p(1)$$

$$=\frac{1}{G_0(\tilde{T}_i)}\frac{1}{\sqrt{n}}\sum_{j=1}^n\int_0^\tau I(\tilde{T}_i \ge s)\frac{dM_j^C(s)}{y(s)}+o_p(1),$$

where $y(s) =\lim_{n\to\infty} (1/n)\sum_{i=1}^n I\left(\tilde{T}_i \ge s\right), M_i^C(t) = (1 - \delta_i) I\left(\tilde{T}_i \le t\right) - \int_0^t I\left(\tilde{T}_i \ge s\right)d\Lambda_C(s)$, and $\Lambda_C(\cdot)$ is the cumulative hazard function of the censoring time C. This leads to

$$Q(\widehat{G}, \theta_0+\frac{u}{\sqrt{n}}) - Q(G_0,\theta_0+\frac{u}{\sqrt{n}})$$

$$=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{G_0(\tilde{T}_i)}|\varepsilon_i - X_i'\frac{u}{\sqrt{n}}|\sum_{j=1}^n\int_0^\tau\frac{I(\tilde{T}_i \ge s)}{y(s)}dM_j^C(s)+o_p(1).$$

Similarly, we have

$$Q(\widehat{G}, \theta_0) - Q(G_0,\theta_0)=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{G_0(\tilde{T}_i)}|\varepsilon_i|\sum_{j=1}^n\int_0^\tau\frac{I(\tilde{T}_i \ge s)}{y(s)}dM_j^C(s)+o_p(1).$$

Therefore,

$$Q(\widehat{G}, \theta_0 + \tfrac{u}{\sqrt{n}}) - Q(G_0, \theta_0 + \tfrac{u}{\sqrt{n}}) - \{Q(\widehat{G}, \theta_0) - Q(G_0, \theta_0)\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{G_0(\tilde{T}_i)}(|\varepsilon_i - X_i' \frac{u}{\sqrt{n}}| - |\varepsilon_i|) \int_0^\tau \sum_{j=1}^{n} \frac{I(\tilde{T}_i \geq s)}{y(s)} dM_j^C(s) + o_p(1).$$

By the similar techniques used for establishing the lower bound of

$Q\left(G_0, \theta_0 + \dfrac{u}{\sqrt{n}}\right) - Q(G_0, \theta_0)$, we can show that the right-hand side of the above expression can be represented as

$$\frac{u'}{\sqrt{n}} \sum_{j=1}^{n} \int_0^\tau \frac{h(s)}{y(s)} dM_j^C(s) + o_p(1),$$

where $h(s) = \lim_{n\to\infty} \dfrac{1}{n} \sum_{i=1}^{n} \dfrac{\delta_i I(\tilde{T}_i \geq s)}{G_0(\tilde{T}_i)} X_i \{I(\varepsilon_i < 0) - I(\varepsilon_i > 0)\}$ is a bounded function on $[0,$

$\tau]$. Then by the martingale central limit theorem, we have that $\dfrac{u'}{\sqrt{n}} \sum_{j=1}^{n} \int_0^\tau \dfrac{h(s)}{y(s)} dM_j^C(s)$ converges in distribution to $u'W_2$, where W_2 is a $(p+1)$-dimensional normal with mean 0 and

variance-covariance matrix $\Sigma_2 = \int_0^\tau \dfrac{h^2(s)}{y(s)} d\Lambda_c(s)$. In summary, we showed that

$$D_n(u) \geq \frac{1}{2}u'\Sigma u + u'(W_1 + W_2) - \kappa_1 O_p(\|u\|) + o_p(1).$$

For the right-hand side in the above expression, the first term dominates the remain terms if $M = \|u\|$ is large enough. So for any $\eta > 0$, as n gets large, we have

$$P\left\{ \inf_{\|u\|=M} D_n(u) > 0 \right\} \geq 1 - \eta,$$

which implies that $\hat{\theta}$ is $\sqrt{n}$-Consistency.

Proof of Theorem 2

(i) Proof of selection-consistency. We will first take the derivative of $Q(\hat{G}, \theta)$ with respect to β_j for $j = q+1, \ldots, p$, at any differentiable point $\theta = \left(\alpha, \beta_a', \beta_b'\right)'$. Then we will show that for

$\sqrt{n}\|\alpha - \alpha_0\| \leq M$, $\sqrt{n}\|\beta_a - \beta_{a,0}\| \leq M$, and $\|\beta_b - \beta_{b,0}\| = \|\beta_b\| \leq \dfrac{M}{\sqrt{n}} \equiv \varepsilon_n$, when n is large,

$\dfrac{1}{\sqrt{n}} \dfrac{\delta}{\delta\beta_j} Q\left(\hat{G}, \theta\right)$, for $j = q+1, \ldots, p$, is negative if $-\varepsilon_n < \beta_j < 0$ and positive if $0 < \beta_j < \varepsilon_n$. Since $Q(\hat{G}, \theta)$ is a piecewise linear function of θ, it achieve its minimum at some breaking point.

Moreover, based on Theorem 1, the minimizer $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta_a'}, \widehat{\beta_b'})'$ of $Q(\hat{G}, \theta)$ is $\sqrt{n}$-Consistency. Thus, each component of $\hat{\beta}_b$ must be contained in the interval $(-\varepsilon_n, \varepsilon_n)$ for all large n. Then as $n \to \infty$, $P(\hat{\beta}_b = 0) \to 1$.

To do this, we have, for $j = q + 1, \ldots, p,$

$$
\frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta_j} Q(\widehat{G}, \theta)
$$

$$
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij} \frac{\delta_i}{\widehat{G}(\tilde{T}_i)} \mathrm{sign}\{\log(\tilde{T}_i) - X_i' \theta\} + \frac{\sqrt{n} \lambda \, \mathrm{sign}(\beta_j)}{|\tilde{\beta}_j|}
$$

$$
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij} \frac{\delta_i}{G_0(\tilde{T}_i)} \mathrm{sign}\{\log(\tilde{T}_i) - X_i' \theta\} - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \int_0^{\tau} \left[\frac{1}{n} \sum_{i=1}^{n} Z_{ij} \frac{\delta_i}{G_0(\tilde{T}_i)} I(\tilde{T}_i \geq s) \mathrm{sign}\{\log(\tilde{T}_i) - X_i' \theta\} \right] \frac{dM_k^C(s)}{y(s)} + \sqrt{n} \lambda \frac{\mathrm{sign}(\beta_j)}{|\tilde{\beta}_j|} + o_p(1).
$$

$$(8)$$

Let $\Delta = \sqrt{n}(\theta - \theta_0)$ and define

$$
V_n(\Delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \frac{\delta_i}{G_0(\tilde{T}_i)} \mathrm{sign}(\varepsilon_i - X_i' \Delta / \sqrt{n}).
$$

Write $V_n(\Delta) = \{V_{n,0}(\Delta), V_{n,1}(\Delta), \ldots, V_{n,p}(\Delta)\}'$. Then the first term at the right-hand side of (8) can be rewritten as $V_{n,j}(\Delta)$. As shown in Theorem 1, $V_n(0)$ converges in distribution to a $(p + 1)$-dimensional normal with mean 0 and variance-covariance matrix Σ_1. Furthermore, following the similar derivations of Koenker and Zhao (1996), we can show that $\sup_{\|\Delta\| \leq M} \|V_n(\Delta) - V_n(0) + \Sigma_1 \Delta\| = o_p(1)$. This implies that $V_{n,j}(\Delta) = O_p(1)$ since $V_n(0)$ converges in distribution to a normal vector, Σ_1 is finite and Δ is bounded.

Next, since $\sqrt{n} \|\theta - \theta_0\|$ is bounded, by the law of large numbers, we have that

$$
\frac{1}{n} \sum_{i=1}^{n} Z_{ij} \frac{\delta_i}{G_0(\tilde{T}_i)} I(\tilde{T}_i \geq s) \mathrm{sign}\{\log(\tilde{T}_i) - X_i' \theta\} \quad \text{converges to}
$$

$$
b_j(s) \equiv E \left\{ Z_{ij} \frac{\delta_i}{G_0(\tilde{T}_i)} I(\tilde{T}_i \geq s) \mathrm{sign}(\varepsilon_i) \right\}. \text{ Thus, the second term at the right-hand side of (8)}
$$
can be written as

$$
\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \int_0^{\tau} \frac{b_j(s)}{y(s)} dM_k^C(s) + o_p(1),
$$

which is also $O_p(1)$ since it converges to a normal variable with mean 0.

Thus, for $j = q + 1, \ldots, p,$

$$
\frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta_j} Q(\widehat{G}, \theta) = O_p(1) + \frac{n \lambda \, \mathrm{sign}(\beta_j)}{\sqrt{n} |\tilde{\beta}_j|}.
$$

Since the LAD estimator $\tilde{\theta}$ is $\sqrt{n}$–Consistency, we have, for $j = q + 1, \ldots, p, \sqrt{n} |\tilde{\beta}_j| = O_p(1)$.

Then based on the assumption $n\lambda \to \infty$, when n is large, the sign of $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta_j} Q(\widehat{G}, \theta)$ is

determined by the sign of β_j. So as n gets large, $\frac{1}{\sqrt{n}}\frac{\delta}{\delta\beta_j}Q(\widehat{G},\theta)$, for $j = q + 1, \ldots, p$, is negative if $-\varepsilon_n < \beta_j < 0$ and positive if $0 < \beta_j < \varepsilon_n$, which implies $P(\hat{\beta}_b = 0) \to 1$ as $n \to \infty$.

(ii) Proof of asymptotic normality. Based on the results established in Theorem 1 and (i) of Theorem 2, we have the minimizer $\hat{\theta}$ is $\sqrt{n}-$consistent and $P(\hat{\beta}_b = 0) \to 1$ as $n \to \infty$. Thus to derive the asymptotic distribution for the estimators of non-zero coefficients, we only need to establish the asymptotic representation for the following function:

$$
\begin{aligned}
S_n(\widehat{G},\upsilon) &= Q\{\widehat{G}, (\beta'_{a'0}+\upsilon'/\sqrt{n},0')'\} - Q\{\widehat{G}, (\beta'_{a'0},0')'\} \\
&= \sum_{i=1}^{n}\frac{\delta_i}{\widehat{G}(\tilde{T}_i)}(|\varepsilon_i - X'_{a'i}\upsilon/\sqrt{n}| - |\varepsilon_i|) + n\lambda\sum_{j=1}^{q}\frac{|\beta_{j0}+\upsilon_j/\sqrt{n}| - |\beta_{j0}|}{|\tilde{\beta}_j|} \\
&= \sum_{i=1}^{n}\frac{\delta_i}{\widehat{G}(\tilde{T}_i)}(|\varepsilon_i - X'_{a'i}\upsilon/\sqrt{n}| - |\varepsilon_i|) + \sqrt{n}\lambda\sum_{j=1}^{q}\frac{sign|\beta_{j0}|\upsilon_j}{|\tilde{\beta}_j|},
\end{aligned}
$$

where $\beta_{a'0}=\left(\alpha_0,\beta'_{a0}\right)'$ and v is a $(q + 1)$-dimensional vector with bounded norm. Define $X_{a'i} = (1, Z_{i1}, \ldots, Z_{iq})'$, $i = 1, \ldots, n$. Since $\sqrt{n}\lambda \to \lambda_0$ and $\tilde{\beta}_j \to \beta_{j0} \neq 0$ and following the similar derivations as those in the proof of Theorem 1, we can show that

$$
S_n(\widehat{G},\upsilon)=\frac{1}{2}\upsilon'\Sigma_{a'}\upsilon+\frac{\upsilon'}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{G_0(\tilde{T}_i)}X_{a'i}\{I(\varepsilon_i<0) - I(\varepsilon_i>0)\}+\frac{\upsilon'}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau\frac{h_{a'}(s)}{y(s)}dM_i^C(s)+\lambda_0\upsilon'b_1+o_p(1),
$$

(9)

where

$$
\Sigma_{a'}=E\left\{f\left(0|X\right)X_{a'}X'_{a'}\right\},\, h_{a'}(s)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i I\left(\tilde{T}_i \geq s\right)}{G_0\left(\tilde{T}_i\right)}X_{a'i}\{I\left(\varepsilon_i<0\right) - I(\varepsilon_i>0)\},\text{ and } b_1=\left(\frac{sign\,(\beta_1)}{|\beta_1|},\ldots,\frac{sign\,(\beta_q)}{|\beta_q|}\right)
$$

. Define

$$
s_i=\frac{\delta_i}{G_0(\tilde{T}_i)}X_{a'i}\{I(\varepsilon_i<0) - I(\varepsilon_i>0)\}+\int_0^\tau\frac{h_{a'}(s)}{y(s)}dM_i^C(s).
$$

Then by the central limit theorem, $\left(1/\sqrt{n}\right)\sum_{i=1}^{n}s_i$ converges in distribution to a $(q + 1)$-dimensional normal vector $W_{a'}$ with mean 0 and variance-covariance matrix $V_{a'}=E\left(s_1s_1'\right)$. The minimizer of $\frac{1}{2}\upsilon'\Sigma_{a'}\upsilon+\upsilon'W_{a'}+\mathbf{v}'\lambda_0\mathbf{b}_1$ is given by $\upsilon_0=-\Sigma_{a'}^{-1}\left(W_{a'}+\lambda_0b_1\right)$. Moreover, the minimizer of $S_n(\hat{G}, v)$ is given by $\sqrt{n}\left(\hat{\beta}_{a'} - \beta_{a'0}\right)$. By the lemma given in Davis et al. (1992), $\sqrt{n}\left(\hat{\beta}_{a'} - \beta_{a'0}\right)$ converges to v_0 in distribution as $n \to \infty$. Therefore we have that $\sqrt{n}\left(\hat{\beta}_{a'} - \beta_{a'0}\right)$ converges in distribution to $(q + 1)$-dimensional normal vector with mean $-\Sigma_{a'}^{-1}\lambda_0b_1$ and variance-covariance matrix $\Sigma_{a'}^{-1}V_{a'}\Sigma_{a'}^{-1}$.

References

Bang H, Tsiatis AA. Median regression with censored cost data. Biometrics 2002;58:643–649. [PubMed: 12229999]

Breiman L. Heuristics of instability and stabilization in model selection. The Annals of Statistics 1996;24:2350–2383.

Buckley J, James I. Linear regression with censored data. Biometrika 1979;66:429–436.

Cai T, Huang J, Tian L. Regularized estimation for the accelerated failure time model. Biometrics 2009;65:394–404. [PubMed: 18573133]

Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society B 1972;34:187–220.

Cox, DR.; Oakes. Analysis of Survival Data. London: Chapman and Hall; 1984.

Davis R, Knight K, Liu J. M-Estimation for autoregressions with infinite variance. Stochastic Process and Their Applications 1992;40:145–180.

Efron B, Hastie T, Johnstone I, Tibshirani R. Least Angle Regression. Annals of Statistics 2004;32:407–451.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 2001;96:1348–1360.

Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. The Annals of Statistics 2002;30:74–99.

He X, Shi P. Bivariate tensor-product B-splines in a partly linear model. Journal of Multivariate Analysis 1996;58:162–181.

He X, Ng P. COBS: Constrained smoothing via linear programming. Computational Statistics 1999;14:315–337.

He X, Ng P, Portnoy S. Bivariate quantile smoothing splines. Journal of the Royal Statistical Society B 1998;60:537–550.

Jin Z, Lin D, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. Biometrika 2003;90:341–353.

Johnson B. Variable selection in semi-parametric linear regression with censored data. Journal of the Royal Statistical Society B 2008;70:351–370.

Johnson B, Lin D, Zeng D. Penalized estimating functions and variable selection in semi-parametric regression models. Journal of the American Statistical Association 2008;103:672–680. [PubMed: 20376193]

Kalbfleisch, J.; Prentice, R. The Statistical Analysis of Failure Time Data. Hoboken, NJ: Wiley; 1980.

Keles S, van der Laan MJ, Dudoit S. Asymptotically optimal model selection method with right censored outcomes. Bernoulli 2004;10:1011–1037.

Knight K. Limiting distributions for L_1 regression estimators under general conditions. The Annals of Statistics 1998;26:755–770.

Knight K, Fu W. Asymptotics for Lasso-type estimators. The Annals of Statistics 2000;28:1356–1378.

Koenker R. Quantile regression for longitudinal data. Journal of Multivariate Analysis 2004;91:74–89.

Koenker R, Geling L. Reappraising medfly longevity: A quantile regression survival analysis. Journal of the American Statistical Society 2001;96:458–468.

Koenker R, Mizera I. Penalized triograms: total variation regularization for bivariate smoothing. Journal of the Royal Statistical Society B 2004;66:145–163.

Koenker R, Ng P, Portnoy S. Quantile smoothing splines. Biometrika 1994;81:673–680.

Koenker R, D'Orey V. Computing regression quantiles. Applied Statistics 1987;36:383–393.

Li H, Luan Y. Boosting proportional hazards models using smoothing splines with applications to high-dimensional microarray data. Bioinformatics 2005;21:2403–2409. [PubMed: 15713732]

Li Y, Zhu J. L1-norm quantile regression. Journal of Computational and Graphical Statistics 2008;17:163–185.

Lu W, Zhang H. Variable selection for proportional odds model. Statistics in Medicine 2007;26:3771–3781. [PubMed: 17266170]

Peng L, Huang Y. Survival analysis with quantile regression models. Journal of the American Statistical Association 2008;103:637–649.

Portnoy S. Censored regression quantiles. Journal of the American Statistical Association 2003;98

Prentice RL. Linear rank tests with right censored data. Biometrika 1978;65:167–179.

Reid N. A conversation with Sir David Cox. Statistical Science 1994;9:439–455.

Ritov Y. Estimation in a linear regression model with censored data. Annals of Statistics 1990;18:303–328.

Rosenwald A, Wright G, Chan W, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. New England Journal of Medicine 2002;346:1937–1947. [PubMed: 12075054]

Therneau, T.; Grambsch, P. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag Inc.; 2000.

Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B 1996;58:267–288.

Tibshirani R. The Lasso method for variable selection in the Cox model. Statistics in Medicine 1997;16:385–395. [PubMed: 9044528]

Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. Annals of Statistics 1990;18:354–372.

Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection via the LAD-LASSO. Journal of Business and Economics Statistics 2007;20:347–355.

Wang H, Wang L. Locally weighted censored quantile regression. Journal of American Statistical Association. 2009 in press.

Ying Z, Jung S, Wei L. Survival analysis with median regression models. Journal of the American Statistical Association 1995;90:178–184.

Ying Z. A large sample study of rank estimation for censored regression data. Annals of Statistics 1993;21:76–99.

Zhang H, Lu W. Adaptive Lasso for Cox's proportional hazards model. Biometrika 2007;94:1–13.

Zhou L. A simple censored median regression estimator. Statistica Sinica 2006;16:1043–1058.

Zou H. The adaptive lasso and its oracle properties. Journal of American Statistical Association 2006;101:1418–1429.

Zou H. A note on path-based variable selection in the penalized proportional hazards model. Biometrika 2008;95:241–247.

Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics 2008;36:1509–1533. [PubMed: 19823597]
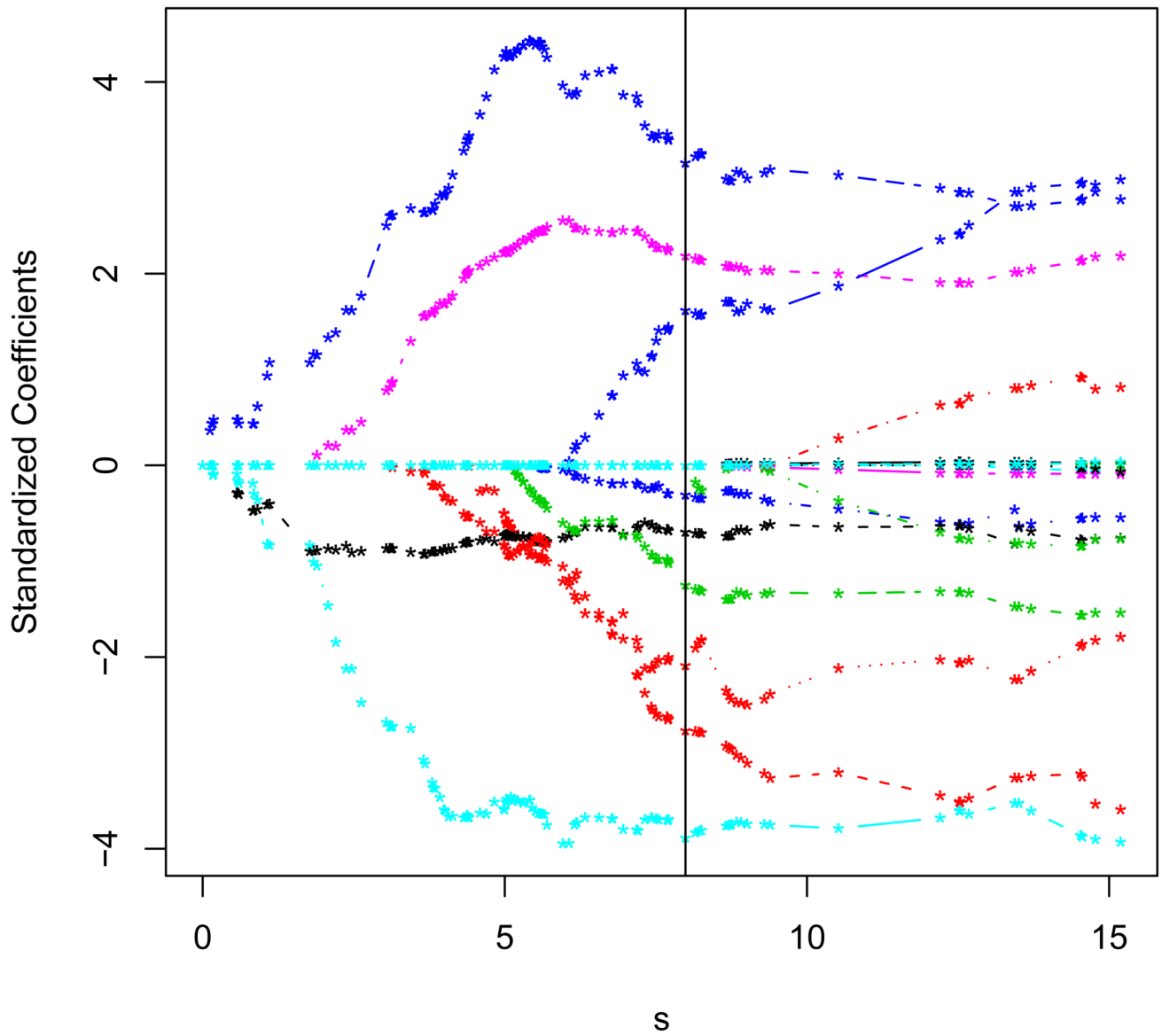
Figure 1.
The solution path of our SCMR-A estimator for PBC data. The solid vertical line denotes the resulting estimator tuned with the proposed BIC criterion.
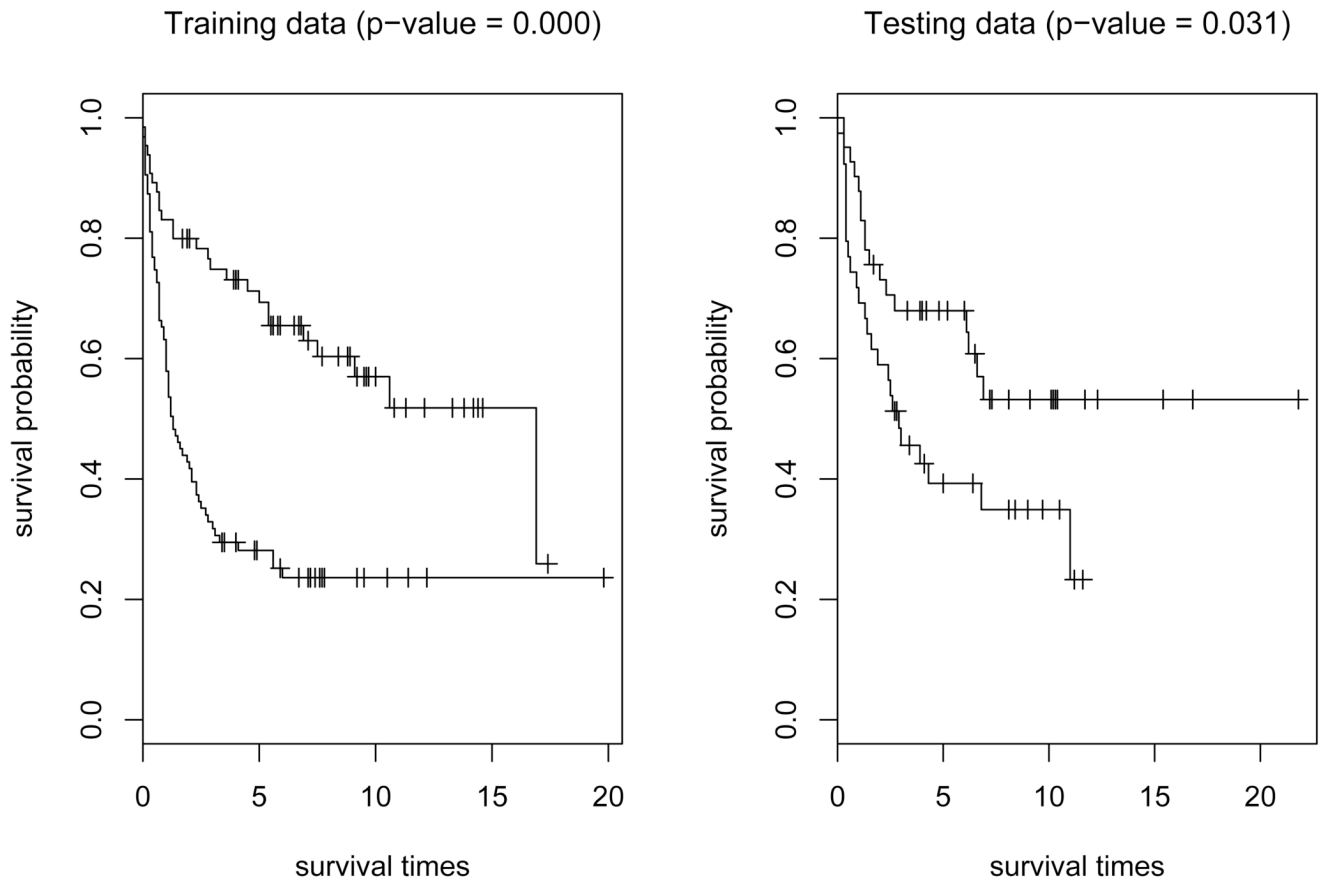
Figure 2.
Kaplan-Meier estimates of survival curves for high-risk and low-risk groups of patients using the selected genes by the SCMR-A.

Table 1

Variable selection results for t(5) error distribution

Prop.	n	Methods	MAD	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	SF	Inc.	Cor.
15%	100	Full	0.39	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.36	92	100	100	100	13	13	26	10	51	0.08	3.38
		SCMR-L	0.41	98	100	100	100	35	42	46	37	16	0.02	2.40
		SCMR-S	0.35	97	100	100	100	15	14	16	15	57	0.03	3.40
		Naive	0.46	90	100	100	100	5	9	15	14	63	0.10	3.57
		Oracle	0.31	100	100	100	100	0	0	0	0	100	0	4
	200	Full	0.28	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.25	100	100	100	100	9	7	8	9	72	0	3.67
		SCMR-L	0.31	100	100	100	100	40	29	36	36	22	0	2.59
		SCMR-S	0.24	100	100	100	100	9	8	10	10	72	0	3.63
		Naive	0.38	100	100	100	100	5	9	12	9	69	0	3.65
		Oracle	0.23	100	100	100	100	0	0	0	0	100	0	4
30%	100	Full	0.50	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.48	87	100	100	100	17	21	26	12	41	0.13	3.24
		SCMR-L	0.56	93	100	100	100	47	49	52	46	8	0.07	2.06
		SCMR-S	0.47	85	100	100	100	12	14	16	19	50	0.15	3.39
		Naive	0.84	79	100	100	100	10	6	12	21	49	0.21	3.51
		Oracle	0.40	100	100	100	100	0	0	0	0	100	0	4
	200	Full	0.41	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.40	99	100	100	100	13	10	13	13	67	0.01	3.51
		SCMR-L	0.47	100	100	100	100	39	46	36	41	18	0	2.38
		SCMR-S	0.36	95	100	100	100	7	9	17	12	63	0.05	3.55
		Naive	0.78	97	100	100	100	5	8	11	12	69	0.03	3.64
		Oracle	0.36	100	100	100	100	0	0	0	0	100	0	4

Prop.: censoring proportion; MAD: mean absolute deviation across 100 runs; SF: number of times the true model is chosen among 100 runs; Inc. and Cor.: mean number of incorrect and correct 0 selected respectively.

Table 2

Estimation of non-zero coefficients for t(5) error distribution

Prop.	n	Methods	β₁ = 0.5			β₂ = 1			β₃ = 1.5		
			Bias	SD	SE	Bias	SD	SE	Bias	SD	SE
15%	100	Full	0.016	0.154	0.195	−0.056	0.158	0.195	−0.069	0.189	0.193
		SCMR-A	−0.052	0.184	0.196	−0.084	0.155	0.193	−0.098	0.186	0.189
		SCMR-L	−0.074	0.160	0.180	−0.144	0.164	0.187	−0.166	0.187	0.188
		SCMR-S	−0.027	0.183	0.217	−0.059	0.165	0.209	−0.074	0.183	0.201
		Naive	−0.126	0.188	0.185	−0.175	0.155	0.184	−0.222	0.174	0.181
		Oracle	0.007	0.151	0.178	−0.062	0.156	0.177	−0.065	0.181	0.177
	200	Full	−0.015	0.109	0.122	−0.038	0.102	0.124	−0.063	0.120	0.120
		SCMR-A	−0.041	0.112	0.127	−0.050	0.104	0.123	−0.078	0.121	0.118
		SCMR-L	−0.065	0.124	0.120	−0.095	0.113	0.122	−0.126	0.130	0.118
		SCMR-S	−0.020	0.116	0.141	−0.037	0.102	0.129	−0.063	0.122	0.131
		Naive	−0.077	0.119	0.117	−0.135	0.098	0.115	−0.193	0.102	0.114
		Oracle	−0.015	0.113	0.116	−0.039	0.098	0.119	−0.065	0.119	0.116
30%	100	Full	−0.054	0.206	0.237	−0.094	0.201	0.232	−0.134	0.220	0.237
		SCMR-A	−0.137	0.222	0.233	−0.137	0.218	0.229	−0.185	0.212	0.231
		SCMR-L	−0.125	0.201	0.214	−0.185	0.224	0.222	−0.254	0.223	0.230
		SCMR-S	−0.102	0.233	0.252	−0.098	0.212	0.261	−0.15	0.205	0.258
		Naive	−0.196	0.206	0.186	−0.33	0.172	0.198	−0.437	0.183	0.196
		Oracle	−0.055	0.190	0.212	−0.09	0.206	0.209	−0.156	0.190	0.214
	200	Full	−0.022	0.134	0.151	−0.092	0.143	0.155	−0.137	0.146	0.145
		SCMR-A	−0.073	0.163	0.156	−0.105	0.144	0.154	−0.15	0.143	0.143
		SCMR-L	−0.081	0.144	0.147	−0.153	0.145	0.151	−0.215	0.154	0.143
		SCMR-S	−0.054	0.169	0.170	−0.083	0.139	0.169	−0.132	0.137	0.170
		Naive	−0.159	0.128	0.128	−0.292	0.117	0.126	−0.435	0.113	0.129
		Oracle	−0.035	0.141	0.140	−0.084	0.141	0.145	−0.133	0.140	0.139

SD: sample standard deviation of estimates; SE: mean of estimated standard errors computed based on 500 bootstraps.

NIH-PA Author Manuscript

Table 3

Variable selection results for Double Exponential error distribution

Prop.	n	Methods	MAD	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	SF	Inc.	Cor.
15%	100	Full	0.36	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.32	94	100	100	100	13	7	9	12	64	0.06	3.59
		SCMR-L	0.39	96	100	100	100	42	37	39	31	17	0.04	2.51
		SCMR-S	0.31	94	100	100	100	16	9	12	11	62	0.06	3.52
		Naive	0.41	94	100	100	100	12	9	11	7	67	0.06	3.61
		Oracle	0.26	100	100	100	100	0	0	0	0	100	0	4
	200	Full	0.24	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.21	100	100	100	100	6	3	8	7	79	0	3.76
		SCMR-L	0.28	100	100	100	100	32	29	31	39	29	0	2.69
		SCMR-S	0.20	100	100	100	100	6	2	5	5	84	0	3.82
		Naive	0.36	100	100	100	100	7	6	10	11	73	0	3.66
		Oracle	0.19	100	100	100	100	0	0	0	0	100	0	4
30%	100	Full	0.47	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.47	88	100	100	100	16	19	8	15	47	0.12	3.42
		SCMR-L	0.53	92	100	100	100	41	44	44	44	16	0.08	2.27
		SCMR-S	0.44	90	100	100	100	12	20	10	12	51	0.10	3.46
		Naive	0.80	72	100	100	100	14	11	10	12	42	0.28	3.53
		Oracle	0.38	100	100	100	100	0	0	0	0	100	0	4
	200	Full	0.36	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.35	99	100	100	100	8	11	15	9	65	0.01	3.57
		SCMR-L	0.43	99	100	100	100	41	34	39	37	17	0.01	2.49
		SCMR-S	0.33	99	100	100	100	8	9	14	8	68	0.01	3.61
		Naive	0.78	96	100	100	100	8	7	11	9	70	0.04	3.65
		Oracle	0.31	100	100	100	100	0	0	0	0	100	0	4

The notations are the same as in Table 1.

Table 4

Estimation of non-zero coefficients for Double Exponential error distribution

Prop.	n	Methods	$\beta_1 = 0.5$			$\beta_2 = 1$			$\beta_3 = 1.5$		
			Bias	SD	SE	Bias	SD	SE	Bias	SD	SE
15%	100	Full	−0.004	0.144	0.204	−0.043	0.168	0.206	−0.074	0.174	0.204
		SCMR-A	−0.078	0.183	0.206	−0.070	0.169	0.202	−0.083	0.172	0.194
		SCMR-L	−0.091	0.172	0.187	−0.126	0.170	0.196	−0.148	0.187	0.192
		SCMR-S	−0.045	0.193	0.214	−0.047	0.165	0.214	−0.057	0.157	0.199
		Naive	−0.107	0.167	0.176	−0.146	0.146	0.179	−0.184	0.144	0.172
		Oracle	−0.017	0.146	0.180	−0.037	0.150	0.178	−0.053	0.158	0.173
	200	Full	−0.030	0.088	0.118	−0.047	0.094	0.115	−0.063	0.105	0.114
		SCMR-A	−0.057	0.095	0.121	−0.061	0.093	0.112	−0.074	0.104	0.110
		SCMR-L	−0.086	0.093	0.115	−0.109	0.101	0.113	−0.132	0.113	0.110
		SCMR-S	−0.037	0.098	0.141	−0.045	0.091	0.124	−0.064	0.101	0.125
		Naive	−0.083	0.096	0.115	−0.138	0.087	0.108	−0.186	0.095	0.110
		Oracle	−0.027	0.085	0.109	−0.043	0.091	0.105	−0.063	0.100	0.105
30%	100	Full	−0.031	0.168	0.270	−0.078	0.191	0.257	−0.140	0.199	0.271
		SCMR-A	−0.100	0.213	0.265	−0.123	0.190	0.256	−0.192	0.205	0.263
		SCMR-L	−0.109	0.199	0.244	−0.159	0.190	0.245	−0.241	0.217	0.258
		SCMR-S	−0.077	0.215	0.244	−0.076	0.185	0.274	−0.161	0.198	0.272
		Naive	−0.227	0.209	0.176	−0.282	0.17	0.202	−0.415	0.167	0.194
		Oracle	−0.021	0.178	0.237	−0.072	0.184	0.228	−0.143	0.190	0.234
	200	Full	−0.060	0.107	0.152	−0.099	0.111	0.147	−0.112	0.118	0.150
		SCMR-A	−0.095	0.116	0.156	−0.118	0.112	0.144	−0.130	0.125	0.145
		SCMR-L	−0.113	0.116	0.147	−0.165	0.120	0.142	−0.194	0.133	0.144
		SCMR-S	−0.072	0.121	0.173	−0.094	0.111	0.169	−0.115	0.128	0.168
		Naive	−0.170	0.128	0.131	−0.301	0.107	0.127	−0.428	0.118	0.127
		Oracle	−0.054	0.105	0.141	−0.095	0.109	0.135	−0.112	0.123	0.137

The notations are the same as in Table 2.

Table 5

Sensitivity Analysis: variable selection results for t(5) error distribution

Prop.	n	Methods	MAD	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	SF	Inc.	Cor.
15%	100	Full	0.38	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.33	95	100	100	100	10	15	25	9	56	0.05	3.41
		SCMR-L	0.40	98	100	100	100	41	43	42	39	19	0.02	2.35
		Oracle	0.29	100	100	100	100	0	0	0	0	100	0	4
	500	Full	0.20	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.18	100	100	100	100	5	6	8	9	80	0	3.72
		SCMR-L	0.23	100	100	100	100	27	31	24	28	31	0	2.90
		Oracle	0.17	100	100	100	100	0	0	0	0	100	0	4
50%	100	Full	0.76	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.79	84	98	99	100	20	18	19	26	30	0.19	3.17
		SCMR-L	0.90	90	99	99	100	44	43	42	51	11	0.12	2.20
		Oracle	0.71	100	100	100	100	0	0	0	0	100	0	4
	500	Full	0.60	100	100	100	100	100	100	100	100	0	0	0
		SCMR-A	0.61	100	100	100	100	13	14	8	14	62	0	3.51
		SCMR-L	0.69	100	100	100	100	47	39	50	47	12	0	2.17
		Oracle	0.58	100	100	100	100	0	0	0	0	100	0	4

The notations are the same as in Table 1.

Table 6

Sensitivity Analysis: estimation of non-zero coefficients for t(5) error distribution

Prop.	n	Methods	$\beta_1 = 0.5$				$\beta_2 = 1$				$\beta_3 = 1.5$			
			Bias	SD	SE	Bias	SD	SE	Bias	SD	SE	Bias	SD	SE
15%	100	Full	0.017	0.187	0.194	−0.047	0.145	0.196	−0.067	0.142	0.197			
		SCMR-A	−0.047	0.195	0.196	−0.076	0.150	0.197	−0.097	0.156	0.193			
		SCMR-L	−0.055	0.188	0.182	−0.141	0.156	0.193	−0.148	0.159	0.191			
		Oracle	0.003	0.168	0.178	−0.037	0.143	0.182	−0.072	0.146	0.182			
	500	Full	0.011	0.076	0.072	−0.035	0.078	0.072	−0.070	0.067	0.073			
		SCMR-A	−0.003	0.078	0.073	−0.044	0.079	0.071	−0.074	0.072	0.072			
		SCMR-L	−0.027	0.081	0.071	−0.087	0.082	0.072	−0.123	0.075	0.072			
		Oracle	0.014	0.077	0.070	−0.034	0.079	0.070	−0.067	0.072	0.071			
50%	100	Full	−0.013	0.238	0.342	−0.167	0.216	0.346	−0.261	0.247	0.348			
		SCMR-A	−0.058	0.261	0.328	−0.229	0.244	0.344	−0.316	0.281	0.342			
		SCMR-L	−0.064	0.234	0.302	−0.272	0.219	0.323	−0.395	0.294	0.326			
		Oracle	0.025	0.237	0.324	−0.166	0.230	0.323	−0.284	0.271	0.312			
	500	Full	0.020	0.135	0.139	−0.182	0.110	0.141	−0.249	0.131	0.141			
		SCMR-A	0.004	0.144	0.142	−0.206	0.128	0.140	−0.275	0.132	0.139			
		SCMR-L	−0.010	0.144	0.135	−0.242	0.119	0.138	−0.322	0.141	0.139			
		Oracle	0.023	0.136	0.134	−0.181	0.119	0.134	−0.256	0.131	0.136			

The notations are the same as in Table 2.

Table 7

Estimation and variable selection for PBC data with censored median regression.

	Full	**SCMR-A**	**SCMR-L**	**SCMR-S**
Intercept	7.62 (0.44)	7.72 (0.40)	7.83 (0.45)	7.71 (0.42)
trt	0.04 (0.14)	0 (−)	0 (−)	0 (−)
age	−3.29 (1.50)	−2.77 (1.36)	−0.93 (0.94)	−2.49 (1.43)
sex	0.03 (0.28)	0 (−)	0.04 (0.24)	0 (−)
asc	−0.57 (0.83)	−0.31 (0.83)	−0.31 (0.81)	0 (−)
hep	−0.05 (0.17)	0 (−)	0.02 (0.17)	0 (−)
spid	−0.09 (0.20)	0 (−)	0 (−)	0 (−)
oed	−0.75 (0.63)	−0.70 (0.61)	−0.70 (0.65)	−0.96 (0.45)
bil	−1.71 (3.24)	−2.09 (3.25)	−1.56 (2.43)	−2.79 (3.68)
chol	−0.87 (3.50)	0 (−)	−0.64 (1.23)	−0.67 (3.07)
alb	2.96 (1.33)	3.16 (1.28)	3.54 (1.17)	2.70 (1.42)
cop	−4.00 (1.97)	−3.89 (1.87)	−2.61 (1.57)	−3.53 (2.02)
alk	2.16 (1.02)	2.19 (0.97)	2.11 (0.87)	2.31 (1.42)
sgot	−0.20 (1.84)	0 (−)	0 (−)	0 (−)
trig	1.16 (2.11)	0 (−)	0 (−)	0.37 (1.74)
plat	−1.61 (1.11)	−1.25 (0.90)	−1.12 (0.67)	−1.48 (0.92)
prot	2.58 (2.31)	1.62 (2.23)	0.93 (1.85)	2.17 (2.72)
stage	0.03 (0.10)	0 (−)	−0.05 (0.09)	0 (−)