Statistics and Its Interface Volume 15 (2022) 303–311

Variable selection for time-varying effects based on interval-censored failure time data

Kaiyi Chen and Jianguo Sun*

Variable selection has recently attracted a great deal of attention and correspondingly, many methods have been proposed. In this paper, we discuss the topic when one faces interval-censored failure time data arising from a model with time-varying coefficients, for which there does not seem to exist a method. For the situation, in addition to identifying important variables or covariates, a desired feature of a variable selection method is to distinguish time-varying coefficients from time-independent ones, which also presents an additional challenge. To address these, a penalized maximum likelihood procedure is presented and in the proposed method, the adaptive group Lasso penalty function and B-spline functions are used. The approach can simultaneously select between time-dependent and time-independent covariate effects. To implement the proposed procedure, an EM algorithm is developed, and a simulation study is conducted and suggests that the proposed method works well in practical situations. Finally it is applied a set of real data on Alzheimer's disease that motivated this study.

Keywords and phrases: Adaptive Group LASSO, Cox model, Interval-censored data.

1. INTRODUCTION

This paper discusses variable or covariate selection when one faces interval-censored failure time data arising from the Cox or proportional hazards (PH) model with time-varying coefficients ([3]). It is well-known that although the standard PH model is the most commonly used regression model for failure time data, it has some disadvantages or limitations. One way to relax these limitations is to allow time-varying coefficients. An example where this is needed is given by a medical study on a treatment whose effect may take time to kick in and also decrease with time due to drug resistance as with most drugs. It is apparent to be crucial to identify when and how fast the treatment becomes effective and ineffective in order to determine optimal treatment strategies among other purposes ([15]).

The analysis of interval-censored failure time data has recently attracted a great deal of attention ([2]; [14]). By interval-censored data, we usually mean that the failure time of interest is observed only to belong to an interval instead of observed exactly. It is easy to see that interval-censored data include right-censored data as a special case. Many authors have discussed variable selection for right-censored data under the standard PH model with constant coefficients ([4]; [16]). Several methods have also been developed for variable selection under the same model but with interval-censored data ([13]; [19]; [9]). In particular, [24] proposed a broken adaptive ridge (BAR) regression approach, and [18] generalized the BAR approach to high-dimensional situations and developed a coordinate-wise optimization algorithm.

Many authors have investigated the analysis of failure time data with time-varying coefficients ([1]; [11]; [15]). There also exists limited literature on variable selection for failure time data with time-varying coefficients ([8]; [22]). In particular, [6] and [20] investigated the problem under the PH model. The former considered the use of the adaptive group Lasso and group SCAD penalty functions and discussed the L_2 convergence rate of the proposed estimators as well as the sparsity and oracle properties. The latter developed an adaptive group LASSO method that can not only identify important variables but also separate the variables with time-varying and time-independent effects. In other words, in addition to selecting non-zero or relevant coefficients, their approach can simultaneous select both time-varying and constant coefficients. However, all methods mentioned above apply only to right-censored data and there is no method available for interval-censored data. In the following, we will generalize the approach given in [20] to interval-censored failure time data.

The rest of this paper is organized as follows. In Section 2, after introducing some notation and the model to be used throughout the paper, the proposed variable selection approach will be described. In the proposed penalized maximum likelihood procedure, the adaptive group LASSO penalty function will be employed and also we will use B-spline functions to approximate time-varying coefficient functions. To implement the proposed procedure, an EM algorithm will be developed in Section 3. Section 4 provides some results obtained from an extensive simulation study conducted to assess the finite sample performance and they suggest that the proposed procedure works well in practice. In Section 5, we apply it to the interval-censored data arising from the Alzheimer's Disease Neuroimaging Initiative that motivated this study, and Section 6 contains some concluding remarks.

*Corresponding author.

2. PENALIZED MAXIMUM LIKELIHOOD VARIABLE SELECTION METHOD

Consider a failure time study that consists of n independent subjects and let T_i denote the failure time of interest associated with subject i. Suppose that for subject i, there exists a vector of covariates denoted by $X_i = (X_{i1}, .., X_{ip})^T$, and one only observes interval-censored data given by $\mathcal{D} = \{(L_i, R_i], X_i\}_{i=1}^n$, where $(L_i, R_i]$ denotes the observed or censored interval to which true failure time T_i belongs. It is obviously that $L_i = 0$ or $R_i = \infty$ corresponds to a left-censored or right-censored observation. In the following, we will assume that the censoring mechanism behind censoring intervals is independent of the failure time of interest. That is, we have independent or non-informative interval censoring ([14]).

To describe covariate effects, we will assume that the failure time of interest T_i follows the Cox PH model given by

$$\Lambda(t|X_i) = \Lambda_0(t)\, \exp[X_i^T \beta(t)] \tag{1}$$

in terms of the cumulative hazard function. In the above, Λ_0 denotes an unknown baseline cumulative hazard function and $\beta(t)$ is a p-dimensional vector of time-varying coefficients. Define a set of B-spline basis functions $\{\, B_l(t),\ l = 1, \ldots, q-1,\ q > 1\,\}$ with $q-1$ degrees of freedom without intercept on a prespecified time interval $[0, \tau]$. In the following, suppose that $\beta(t)$ can be approximated or written as $\beta(t) = \Theta F(t)$, where Θ is an unknown $p \times q$ matrix of parameters to be estimated and $F(t) = \{1, B_1(t), ..., B_{q-1}(t)\}^T$. It follows that one can write each unknown time varying coefficient as $\beta_j(t) = \Theta_j F(t)$, where Θ_j denotes the jth row of Θ, $j = 1, \ldots, p$.

Define $\theta = vech(\Theta)$, the vectorization of the matrix Θ by row, and also define $\tilde{X}_i^T = (X_{i1}, ..., X_{i1}, ..., X_{ip}, ...X_{ip})_{1\times pq}$ and

$$\tilde{F}(s) = (1, B_1(s), ...B_{q-1}(s), ..., 1, B_1(s), ...B_{q-1}(s))_{1\times pq}\,,$$

the corresponding design matrix and B-spline vector, respectively. Then the observed likelihood function has the form

$$\begin{aligned} L_n(\theta, \Lambda_0) &= \prod_{i=1}^n \left\{ \exp\left[-\int_0^{L_i} e^{X_i^T \Theta F(s)} d\Lambda_0(s) \right] \right. \\ &\quad \left. - \exp\left[-\int_0^{R_i} e^{X_i^T \Theta F(s)} d\Lambda_0(s) \right] \right\} \\ &= \prod_{i=1}^n \left\{ \exp\left[-\int_0^{L_i} e^{\tilde{X}_i^T \odot \tilde{F}(s) \theta} d\Lambda_0(s) \right] \right. \\ &\quad \left. - \exp\left[-\int_0^{R_i} e^{\tilde{X}_i^T \odot \tilde{F}(s) \theta} d\Lambda_0(s) \right] \right\}, \end{aligned} \tag{2}$$

where $\odot$ denotes the component-wise multiplication. For the variable or covariate selection, we propose to minimize the negative penalized log likelihood function

$$Q_{\alpha_n}(\theta, \Lambda_0) = -l_n(\theta, \Lambda_0) + P(\theta; \alpha_n)\,, \tag{3}$$

where $l_n(\theta, \Lambda_0) = \log[L_n(\theta, \Lambda_0)]$ and $P(\theta; \alpha_n)$ denotes a penalty function that depends on the tuning parameter α_n.

Let θ_j denote the part of θ corresponding to the jth component of the covariates, $j = 1, ..., p$. In the following, we will consider two adaptive group LASSO (AGLasso) penalty functions, the combined and separate penalty functions. The former has the form

$$P(\theta; \alpha_n) = \alpha_n \sum_{j=1}^p \Omega_j ||\theta_j|| = \alpha_n \sum_{j=1}^p \Omega_j ||\Theta_j||\,, \tag{4}$$

where Ω_j denotes a weight for group j, which will be taken to be

$$\Omega_j = \frac{\sqrt{q}}{||\tilde{\theta_j}||} \tag{5}$$

by following [21] and [23]. Here $\tilde{\theta_j}$ denotes an initial, consistent estimator of θ_j that will be discussed below. It is easy to see that the weight above tends to put more penalty on the smaller norm of $||\tilde{\theta_j}||$, and this penalty treats each row in Θ as a single group without distinguishing whether β_j is time-varying or not.

To describe the separate penalty function, write $\Theta_j = (\Theta_{j,1}, \Theta_{j,-1})$. Here $\Theta_{j,1}$ denotes the component corresponding to the time-independent intercept (overall effect) or the coefficient of the first component, one, in $F(t)$ and $\Theta_{j,-1}$ the rest of parameters corresponding $\{B_1(t), B_2(t), ..., B_{q-1}(t)\}$ in $F(t)$. Then the separate penalty function is given by

$$P(\theta; \alpha_n) = \alpha_n \sum_{j=1}^p \{\Omega_{j1}|\Theta_{j,1}| + \Omega_{j2}||\Theta_{j,-1}||\}\,, \tag{6}$$

where Ω_{j1} and Ω_{j2} denote the weights that can be calculated as in (5). Here we break down each $\beta_j(t)$ into two parts, and if a covariate coefficient is picked out to be nonzero, it can separate between time-varying coefficients and time-independent coefficients.

3. PENALIZED EM ALGORITHM

Now we discuss the minimization of $Q_{\alpha_n}(\theta, \Lambda_0)$ given in (3), which is not straightforward partly due to the involvement of the unknown function Λ_0. For this, we will develop a penalized EM algorithm by using the technique discussed in [17] among others. In particular, we will adopt the nonparametric approach by treating Λ_0 as a step function with nonnegative jumps at the endpoints of the smallest intervals that bracket the failure times of interest.

Let $0 = t_0 < t_1 < ..., < t_m$ denote the sequence of time points consisting of zero and the unique values of $\{\, L_i > 0,\ R_i < \infty,\ 1 = 1, ..., n\,\}$ and suppose that Λ_0 is a step function with jump size λ_k at t_k with $\lambda_k = 0$. Then the likelihood function $L_n(\theta, \Lambda_0)$ can be rewritten as

$$
\begin{aligned}
L_n(\theta,\Lambda_0) =& \prod_{i=1}^{n}\Big\{exp\Big[-\sum_{t_k<L_i} e^{\tilde{X}_i^T\odot\tilde{F}(s)\theta}\lambda_k\Big] \\
& - exp\Big[-\sum_{t_k\le R_i} e^{\tilde{X}_i^T\odot\tilde{F}(s)\theta}\lambda_k\Big]\Big\} \\
=& \prod_{i=1}^{n}\Big\{exp\Big[-\sum_{t_k<L_i} e^{\tilde{X}_i^T\odot\tilde{F}(s)\theta}\lambda_k\Big]\Big\} \\
& \Big\{1-exp\Big[-\sum_{L_i<t_k\le R_i} e^{\tilde{X}_i^T\odot\tilde{F}(s)\theta}\lambda_k\Big]\Big\}^{I(R_i<\infty)} \\
=& \prod_{i=1}^{n}\Big\{exp\Big[-\sum_{t_k<L_i} e^{X_{ik}^T\theta}\lambda_k\Big]\Big\} \\
& \Big\{1-exp\Big[-\sum_{L_i<t_k\le R_i} e^{X_{ik}^T\theta}\lambda_k\Big]\Big\}^{I(R_i<\infty)},
\end{aligned}
\tag{7}
$$

where $X_{ik}^T = \tilde{X}_i^T \odot \tilde{F}(t_k)$.

Let $\{W_{ik},\ \ i = 1, ..., n,\ \ k = 1, ..., m\}$ denote the independent Poisson random variables with the means $e^{X_i^T\ominus F(s)}\lambda_k = e^{X_{ik}^T\theta}\lambda_k$. It is easy to see that the likelihood function above can also be written as

$$
\begin{aligned}
L_n(\theta,\Lambda_0) =& \prod_{i=1}^{n}\Big\{Pr\Big(\sum_{t_k\le L_i} W_{ik}=0\Big)\Big\} \\
& \Big\{Pr\Big(\sum_{L_i<t_k\le R_i} W_{ik}>0\Big)\Big\}^{I(R_i<\infty)} \\
=& \prod_{i=1}^{n}\Big\{Pr\sum_{t_k\le L_i}(W_{ik}=0)\Big\} \\
& \Big\{1-Pr\Big[\sum_{L_i\le t_k\le R_i}(W_{ik}=0)\Big]\Big\}^{I(R_i<\infty)} \\
=& \prod_{i=1}^{n}\Big\{exp\Big[-\sum_{t_k<L_i} e^{X_{ik}^T\theta}\lambda_k\Big]\Big\} \\
& \Big[1-exp\Big[-\sum_{L_i<t_k\le R_i} e^{X_{ik}^T\theta}\lambda_k\Big]\Big]^{I(R_i<\infty)}.
\end{aligned}
\tag{8}
$$

In other words, the maximization of (7) is equivalent to the maximization of the likelihood function above based on the data $(L_i, R_i, X_i, \sum_{t_k\le L_i} W_{ik} = 0, I(R_i < \infty)\sum_{L_i<t_k\le R_i} W_{ik} > 0)$, $i = 1, ..., n$, which motivated the following EM algorithm.

In the EM algorithm, we treat the W_{ik}'s as missing data, which yields the pseudo-complete-data log likelihood function

$$
\begin{aligned}
l_n^* =& \sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k\le R_i^*)\Big[W_{ik}log(\lambda_k exp\{X_{ik}^T\theta\}) \\
& - \lambda_k exp\{X_{ik}^T\theta\} - log(W_{ik}!)\Big],
\end{aligned}
\tag{9}
$$

where $R_i^* = L_i I(R_i = \infty) + R_i I(R_i < \infty)$. Correspondingly, based on the pseudo-complete-data, the objective function given in (3) can be rewritten as

$$
\begin{aligned}
Q_{\alpha_n}^*(\theta) =& -\sum_{i=1}^{n}\sum_{k=1}^{m} I(t_k\le R_i^*)\Big[W_{ik}log(\lambda_k exp\{X_{ik}^T\theta\}) \\
& - \lambda_k exp\{X_{ik}^T\theta\} - log(W_{ik}!)\Big] + P(\theta;\alpha_n).
\end{aligned}
\tag{10}
$$

In the E-step of the EM algorithm, we need to calculate the following expectation

$$
\begin{aligned}
& E(W_{ik}) \\
&= \frac{\lambda_k exp\{X_{ik}^T\boldsymbol{\theta}\} I(L_i<t_k\le R_i, R_i<\infty)}{1-exp\Big\{\sum_{t_k\le L_i}\lambda_k exp\{X_{ik}^T\boldsymbol{\theta}\}-\sum_{t_k\le R_i}\lambda_k exp\{X_{ik}^T\boldsymbol{\theta}\}\Big\}}.
\end{aligned}
\tag{11}
$$

In the M-step of the EM algorithm, with respect to the baseline jump sizes λ_k's, we have the closed-form solution

$$
\lambda_k = \frac{\sum_{i=1}^{n} I(t_k\le R_i^*)\hat{E}(W_{ik})}{\sum_{i=1}^{n} I(t_k\le R_i^*)exp\{X_{ik}^T\boldsymbol{\theta}\}},\quad k = 1, ..., m\,. \tag{12}
$$

With respect to estimation of the parameter θ, by following [20], we will modify the iterative group shooting algorithm (IGSA). More specifically, define

$$
G(\theta|\Lambda_0) = -\frac{\partial l_n^*(\boldsymbol{\theta},\boldsymbol{\Lambda_0})}{\partial\boldsymbol{\theta}}\ ,\ H(\theta|\Lambda_0) = -\frac{\partial^2 l_n^*(\boldsymbol{\theta},\boldsymbol{\Lambda_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\,,
$$

the partial gradient vector and the partial Hessian matrix of θ, respectively. Suppose that the true value $(\tilde{\theta}, \tilde{\Lambda_0})$ satisfies $G(\tilde{\theta}, \tilde{\Lambda_0}) = 0$, and let $\mathbb{X}$ be the pseudo design matrix defined such that $\mathbb{X}^T\mathbb{X}$ is the Cholesky decomposition of H or $H(\theta|\tilde{\Lambda_0}) = \mathbb{X}^T\mathbb{X}$. Also define the pseudo response vector $Y = (\mathbb{X}^T)^{-1}[H(\theta|\tilde{\Lambda_0})\theta - G(\theta|\tilde{\Lambda_0})]$. Then the second-order Taylor expansion of (10) yields the following quadratic form

$$
\frac{1}{2}||Y-\mathbb{X}\theta|| + \alpha_n\sum_{j=1}^{p}\Omega_j||\theta_j||\,, \tag{13}
$$

which transfers the minimization of (10) to a penalized least square problem.

By [21], the necessary and sufficient conditions for θ to be a solution to the penalized least square (13) are

$$-\mathbb{X}_j^T(Y-\mathbb{X}\theta)+\frac{\alpha_j\theta_j}{||\theta_j||}=0,\quad \theta_j\neq 0, \tag{14}$$

$$||-\mathbb{X}_j^T(Y-\mathbb{X}\theta)||\leq \alpha_j,\ \ \theta_j=0, \tag{15}$$

where $\alpha_j=\alpha\,\Omega_j$. Note that the closed-form solution of Yuan and Lin ([21]) is not applicable here because $\mathbb{X}$ is a triangular matrix but not group-orthonormal. The condition (14) is equivalent to

$$S_j=\left(\mathbb{X}_j^T\mathbb{X}_j+\frac{\alpha_j I_q}{||\theta_j||}\right)\theta_j, \tag{16}$$

where $S_j=\mathbb{X}_j^T(Y-\mathbb{X}\theta_{-j})$ with $\theta_{-j}=(\theta_1^T,...,\theta_{j-1}^T,0^T,\theta_{j+1}^T,...,\theta_p^T)^T)^T$. This gives the iterative update of θ_j as

$$\theta_j^{(k)}=\left(\mathbb{X}_j^T\mathbb{X}_j+\frac{\alpha_j I_q}{||\theta_j^{(k-1)}||}\right)^{-1}S_j. \tag{17}$$

Let v be an integer, denoting the number of iterations. Then given the tuning parameter α_n, the developed EM algorithm can be summarized as follows.

- Step 1: Choose the initial estimate $\boldsymbol{\theta}^{(0)}$ as discussed below and set $\lambda_k^{(0)}=1/m,\ k=1,...,m$.
- Step 2: At the v-th step, calculate the conditional expectation $\hat{E}(W_{ik}|\boldsymbol{\theta}^{(v)},\lambda^{(\mathbf{v})})$ given in (11).
- Step 3: Also at the v-th step, update the λ_k's by using (12).
- Step 4: Again at the v-th step, update the first derivative and the second derivative of the log-likelihood as

$$G=\sum_{i=1}^n\sum_{k=1}^m I(t_k\leq R_i^*)\hat{E}(W_{ik}|\boldsymbol{\theta}^{(v)},\lambda^{(\mathbf{v})})\Bigg[\mathbf{X_{ik}} - \frac{\sum_{j=1}^n I(t_k\leq R_j^*)X_{jk}exp\{X_{jk}^T\boldsymbol{\theta}^{(v)}\}}{\sum_{j=1}^n I(t_k\leq R_j^*)exp\{X_{jk}^T\boldsymbol{\theta}^{(v)}\}}\Bigg],$$

$$H=-\sum_{i=1}^n\sum_{k=1}^m I(t_k\leq R_i^*)\hat{E}(W_{ik}|\boldsymbol{\theta}^{(v)},\lambda^{(\mathbf{v})}) \left[\frac{S^{(2)}(\boldsymbol{\theta}^{(v)})}{S^{(0)}(\boldsymbol{\theta}^{(v)})}-\left(\frac{S^{(1)}(\boldsymbol{\theta}^{(v)})}{S^{(0)}(\boldsymbol{\theta}^{(v)})}\right)^{\otimes 2}\right],$$

  where $\otimes$ means an inner product and $S^{(b)}(\boldsymbol{\theta}^{(v)})=\sum_{u=1}^n I(t_k\leq R_u^*)X_{uk}^{\otimes b}exp\{X_{uk}^T\boldsymbol{\theta}^{(v)}\}$, $b=0,1,2$.
- Step 5: Update θ by using (17) based on the modified IGSA.
- Step 6: Repeat Steps 2–5 until the convergence.

For the implementation of the algorithm above, one needs to choose the initial estimate for θ in Step 1 and a convergence criterion in Step 6. For the former, we suggest to use the regular maximum likelihood estimate without the penalty function, which can be obtained by using the R function *survreg()* in the package **survival**, which fits parametric models to interval-censored data. To check the convergence, any criterion could be used and a natural one, which will be used below, is to apply the sum of the absolute differences of the estimates at two successive iterations.

To apply the variable selection procedure above, it is apparent that one also needs to choose q, the degrees of freedom for the B-spline functions, and the tuning parameter α_n. For the former, we suggest to try different values and compare the obtained results, and more comments on this will be given below. On the selection of the tuning parameter α_n, different methods could be used and in the numerical studies below, the generalized cross-validation (GCV) criterion will be employed. More specifically, the grid approach will be used to choose α_n that minimizes

$$GCV(\alpha_n)=\frac{-l_n(\theta)}{n[1-p(\alpha_n)/n]^2}.$$

In the above, $p(\alpha_n)=tr\{H+\alpha_n D)^{-1}H\}$, which can be viewed as the number of effective parameters, where $D=diag\{diag(\Omega_1/||\boldsymbol{\Theta}_1),\ldots,diag(\Omega_p/||\boldsymbol{\Theta}_\mathbf{p})\}$ and H denotes the second derivative of the log likelihood function.

4. A SIMULATION STUDY

An extensive simulation study was conducted to evaluate the finite sample performance of the variable selection procedure proposed in the previous sections. In the study, the covariates were generated from either the Bernoulli distribution with the probability of success 0.5 independently or the multivariate normal distribution with mean zero, variance one and the covariance $0.5^{(j-k)}$ between the jth and kth components, $j,\,k=1,\ldots,p$. For regression coefficients $\beta(t)$, we considered three settings with the non-zero coefficients given as follows

- Setting 1: $\beta_1(t)=t$ and $\beta_2=0.5$;
- Setting 2: $\beta_1(t)=0.5+sin(\frac{\pi t}{2})$ and $\beta_2=0.5$;
- Setting 3: $\beta_1(t)=-1-cos(\pi t)I(0\leq t\leq 1)$, $\beta_2(t)=0.5+sin(\pi/2t)$, $\beta_3=1$ and $\beta_4=-1$.

Let s denote the number of non-zero coefficients. Then we have that $s=2$ for Settings 1 and 2 and $s=4$ for Setting 3.

Given the covariates X_i's and $\beta(t)$, we generated the failure times T_i's by solving the equations $S_i(t|X_i,\beta(t))=u_i$, $i=1,...,n$, where the u_i's were generated independently from the uniform distribution over $(0,1)$ and

$$S_i(t|X_i,\beta(t))=exp\left[-\int_0^t exp(X_i^T\beta(s))d\Lambda_0(s)\right]$$

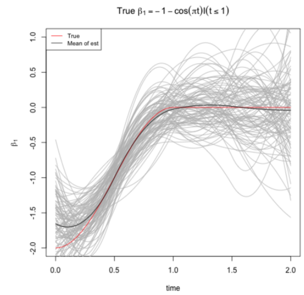
with $\Lambda_0(t)=t$. For the generation of interval-censored data, to mimic clinical studies, we assumed that there exist

Table 1. *Simulation results on the average numbers of true non-zero coefficient (NTP) and false non-zero coefficients (NFP) and MSE*

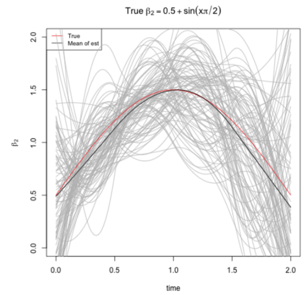
n	p	Setting	NTP	NFP	MSE
400	10	1	1.98	0.02	0.0937
		2	1.93	0.03	0.0621
		3	4	0.23	0.0204
	20	1	2	0.1	0.0941
		2	1.94	0.16	0.0634
		3	4	0.5	0.0201
600	10	1	2	0.03	0.0405
		2	1.98	0.05	0.0163
		3	4	0.25	0.0184
	20	1	2	0.09	0.0779
		2	1.97	0.14	0.0157
		3	4	0.48	0.0143

$M = 10$ or 20 equally spaced examination time points over $(0, \tau)$ and each subject was observed at each of these time points with probability 0.5. Let $\tau = 2$ in all simulation settings. Then for subject i, the observed interval $(L_i, R_i]$ was determined by setting L_i and R_i as the largest, real examination time point that is smaller than T_i and the smallest, real examination time point that is greater than T_i, respectively. The results given below are based on $n = 400$ or 600, $p = 10$ or 20, with 100 replications, and the quadratic B-spline with 5 degree of freedom and equally spaced knots over time $(0, \tau)$ for the non-zero coefficients.

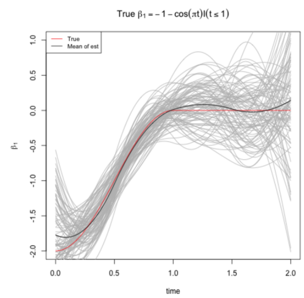
Table 1 presents the results on the covariate selection obtained with the use of the combined penalty function under the three settings for non-zero coefficients with $M = 20$ and discrete covariates. Note that for the situation, it was assumed that we know which coefficients are time-varying and which ones are constant and all zero coefficients are treated to be constants. In the table, we calculated the average number of true non-zero coefficients selected (NTP), the average number of zero coefficients selected (NFP), and the average of the mean square errors (MSE). Here the MSE was determined by the average of $[\hat{\beta}(t) - \beta(t)]^T[\hat{\beta}(t) - \beta(t)]$ over 100 equally spaced time points in $(0, \tau)$. The results suggest that the proposed variable selection approach seems to perform well and especially, as expected, the performance became better when the sample size increased.



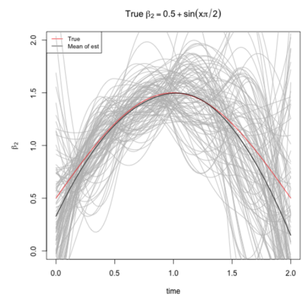
(a) n=400, β_1



(b) n=400, β_2



(c) n=600, $\beta 1$



(d) n=600, β_2

Figure 1. *Estimated curves of the two nonzero time-varying coefficients under Setting 3 with* $p = 10$.

To give more details about the results presented in Table 1, Table 2 shows the frequency of the selection among 100 replications for each of the 10 covariates under the three settings with $p = 10$. Note that for Settings 1 and 2, the first two covariates have non-zero coefficients and for Setting 3, the first four covariates have non-zero coefficients. Figure 1 displays the estimated curves of the two nonzero time-varying coefficients under Setting 3 also with $p = 10$ and with the black solid line representing the mean estimate. For comparison, the true curve is also included in the red solid line. Again these results indicate that the proposed variable selection and estimation method appears to work well.

Note that in Table 1, we set $M = 20$. To see the possible effect of M on the variable selection, we repeated the study giving Table 1 except with $M = 10$ and Table 3 presents the results on the variable selection under Setting 2. Note that

Table 2. *Simulation results on the number of times that each covariate was selected over 100 replications with* $p = 10$

n	Setting	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
400	1	100	98	0	0	1	0	0	0	1	0
	2	94.17	99	0	1	0	0	0	0	1	1
	3	100	100	100	100	3	3	4	3	7	3
600	1	100	100	1	0	1	0	0	1	0	0
	2	97.67	100	2	0	1	0	1	1	0	0
	3	100	100	100	100	5	4	3	5	3	5

Table 3. Simulation results on the average numbers of true non-zero coefficient (NTP) and false non-zero coefficients (NFP) and MSE under Setting 2

		$M = 10$			$M = 20$		
n	p	NTP	NFP	MSE	NTP	NFP	MSE
400	10	1.76	0.01	0.0654	1.93	0.03	0.0621
	20	1.74	0.02	0.0631	1.94	0.16	0.0634
600	10	1.84	0	0.0566	1.98	0.05	0.0163
	20	1.74	0.02	0.0784	1.97	0.14	0.0157

Table 4. Simulation results on the average numbers of true non-zero coefficient (NTP) and false non-zero coefficients (NFP) and MSE with multivariate normal covariates and $p = 20$

n	M	Setting	NTP	NFP	MSE
400	10	1	1.99	0.39	0.1896
		2	1.97	0.37	0.0362
	20	1	2	0.48	0.1027
		2	2	0.52	0.0024
600	10	1	2	0.4	0.1565
		2	1.99	0.23	0.0397
	20	1	2	0.37	0.07626
		2	2	0.37	0.0040

Table 5. Simulation results on the average numbers of true non-zero coefficient (NTP) and false non-zero coefficients (NFP) and MSE with the separate penalty function and $p = 10$

Covariates	n	M	NTP	NFP	MSE
Binary	400	20	1.98	0.01	0.1757
		10	1.99	0.01	0.1815
	600	20	1.96	0.05	0.1863
		10	1.98	0.02	0.1755
Normal	400	20	1.98	0.10	0.1758
		10	2	0.15	0.1772
	600	20	1.98	0.05	0.1831
		10	1.98	0.15	0.1808

a smaller M means less information about the failure time of interest. Table 3 suggests that it does have some effects on NTP but not much on NFP and MSE. Table 4 gives the results on the variable selection obtained under the same set-up as with Table 1 except based on the multivariate normal covariates under Settings 1 and 2 and with $p = 20$. One can see that they gave similar conclusions as Tables 1 and 3 and again indicate that the proposed variable selection procedure works well.

Note that all results given above were obtained with the use of the combined penalty function. Table 5 gives the variable selection results obtained with the use of the separate penalty function and $p = 10$ for both discrete and continues covariates. Here instead of the settings used above, we considered two non-zero coefficients given by

Table 6. Simulation results on the numbers of times that the INT and TV components were selected with the separate penalty function and $p = 10$

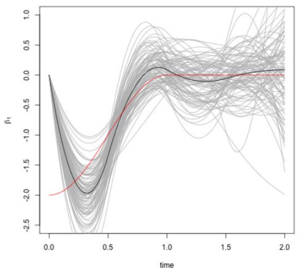
Covariates	n	M	X_1		X_2	
			INT.	TV	INT.	TV
Binary	400	20	0	100	98	10
		10	0	100	96	16
	600	20	0	100	99	14
		10	0	100	98	10
Normal	400	20	0	100	98	10
		10	0	100	98	12
	600	20	0	100	100	4
		10	0	100	99	4

$\beta_1(t) = -1 - cos(\pi t) I(0 \leq t \leq 1)$ and $\beta_2 = 1$. One can see from the table that the results are similar to those given above based on the combined penalty function. Note that for the situation, it was assumed that there is no knowledge about the time dependence of covariate effects and all coefficients were treated to be time-varying. It is expected that the method can identify the constant coefficients from the time-varying ones. To see this, Table 6 reports the frequencies that the intercept (INT.) component and the time varying (TV) component for each of the two nonzero effects were selected. One can see that the first covariate X_1 was correctly selected to have time-varying effects all the time, while the second covariate X_2 was also correctly selected to have constant (INT.) effects over 95%.

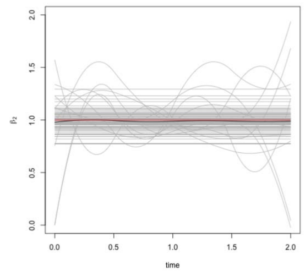
To further see the results given in Table 5, Figure 2 shows the estimated curves of the two nonzero coefficients with the black and red solid lines representing the mean estimate and the true curve, respectively. One can see that the estimated $\beta_1(t)$ is reasonably well except near $t = 0$ as expected and the estimated β_2 is straight lines most of the time as indicated in Table 6. In the simulation study, we also considered other set-ups, including different true $\Lambda_0(t)$ and different values for p, s and M, and obtained similar results.

5. ANALYSIS OF THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

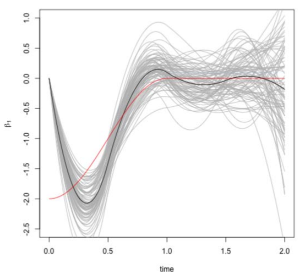
Now we apply the penalized maximum likelihood variable selection procedure proposed in the previous sections to the data arising from the Alzheimer's Disease Neuroimaging



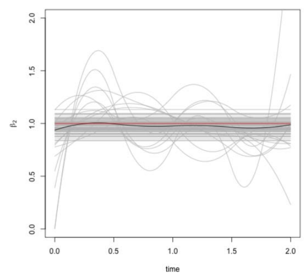
(a) n=400, β_1



(b) n=400, β_2



(c) n=600, $\beta 1$



(d) n=600, β_2

Figure 2. The red line, black line and blue lines are the true curve of coefficient, the mean of estimated curve of coefficient over 100 replicates. The gray lines are 100 replicates. All figures are for normal continues covariates and $M = 20$, $p = 10$.

Initiative (ADNI). It is an on-going multisite longitudinal study with the primary goal being to investigate if various possible risk factors, including demographic factors, clinical factors and biological markers, can be used to measure the progression to Alzheimer's disease (AD). In the study, the participants, recruited from 57 sites in the United States and Canada, are classified into three groups and by following others, we will focus on the subjects in the mild cognitive impairment (MCI) group. Among others, one variable of interest, the focus here, is the time from the baseline visit to the AD conversion. Since the participants are only examined only during their clinical visits, only interval-censored observations are available for the AD conversion time.

For the analysis, by following others, we will consider the 310 MCI subjects with the complete information on 25 demographic, clinical and subject's MRI volumetric data-related risk factors ([12], [5], [10], [7]). The demographic factors include gender (1 for male and 0 for female), race (1 for white and 0 for others), marital status (1 for married and 0 otherwise), age, years of receiving education (PTEDUCAT), and apolipoprotein E genotype (APOE4). The clinical factors consist of the participant's AD Assessment Scale scores of 11 and 13 items (ADAS11 and ADAS13), the delayed word recall score in ADAS (ADASQ4), the clinical dementia rating scale-sum of boxes score (CDRSB), the mini-mental state examination score (MMSE), Rey auditory verbal learning test score of immediate recall (RAVLT.imediate), the learning ability (RAVLT.learning), the total number of words that were forgotten in the RAVLT delayed memory test (RAVLT.forgetting), the percentage of words that were forgotten in the RAVLT delayed memory test (RAVLT.perc.forgetting), the participants' digit symbol substitution test score (DIGITSCOR), the trails B score (TRABSCOR), and the functional assessment questionnaire score (FAQ). The participants' MRI volumetric data-related factors include ventricles, hippocampus, whole brain (WholeBrain), entorhinal, fusiform gyrus (Fusiform), middle temporal gyrus (MidTemp), and intracerebral volume (ICV).

To apply the proposed method, we first used the separate penalty functions and identified three time-varying coefficients corresponding to RAVLT.immediate, FAQ and MidTemp. Then we used the combine penalty function with treating the three coefficients above as time-varying ones and all other as constant coefficients. Table 7 presents the selected covariates by the proposed approach and for comparison, we also obtained and include the selected covariates by the same method but assuming that all covariates had constant effects (ALasso). Here as in the simulation study, the quadratic B-spline with 5 degrees of freedom was used. In the table, in addition to the estimated effects, we also give the estimated standard errors obtained based on the bootstrap procedure with 200 bootstrap samples.

It can be seen from Table 7 that the proposed method selected less covariates than the method that assumed that all covariates had constant effects. However, all of the three covariates with significant effects based on the latter were selected by the proposed procedure, which also suggested that they had time-varying effects. Figures 3–5 display the three estimated time-varying coefficients along with the 95% pointwise confidence bands given by the bootstrap procedure based on 200 bootstrap samples. They suggest that both covariates RAVLT.immediate and MidTemp clearly seem to indeed have significantly negative time-varying effects and the largest effect of RAVLT.immediate occurred earlier than that of MidTemp. We also performed the analysis with the use of the B-spline with other degrees of freedom and obtained similar results.

6. CONCLUDING REMARKS

This paper discussed variable selection for interval-censored failure time data arising from Cox model with time-varying coefficients. For the problem, a penalized maximum likelihood procedure was presented and in the proposed method, the adaptive group Lasso penalty function and B-splines were used. For the implementation of the procedure, an innovated EM algorithm was developed with the use of Poisson random variables in data augmentation. To assess the performance of the proposed approach, an extensive simulation study was conducted and indicates that the

Table 7. The selected factors along with their estimated effects for the ADNI based on the proposed method (AGLasso) and the ALasso approach

Covariates	ALasso		AGLasso	
	Coef.	Std.Err	Coef.	Std.Err
White	0.9314	0.6965	1.0435	2.649
AGE	−0.1954	0.1501	−0.2334	0.7801
APOE4	−0.1954	0.1501	−0.1763	0.2318
DAS13	0.2727	0.9394	0.1291	0.3551
RAVLT.immediate	−0.5024	0.1646	See Figure 3	
RAVLT.perc.forgetting	0.0539	0.2584		
DIGITSCOR	−0.0463	0.1503		
FAQ	0.3404	0.0994	See Figure 4	
Entorhinal	−0.2450	0.1359	−0.1455	0.1393
MidTemp	−0.4816	0.1628	See Figure 5	
ICV	0.1087	0.2889		

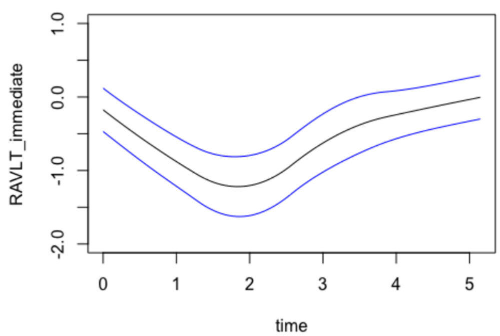


Figure 3. The estimated time-varying effects for RAVLT.immediate.

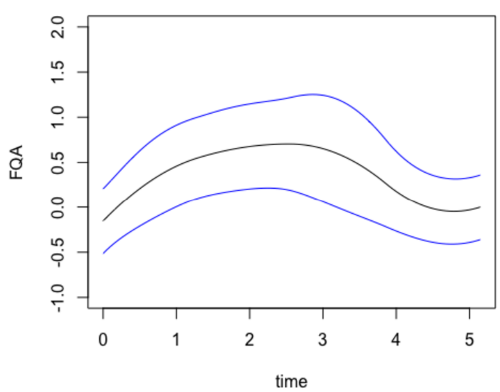


Figure 4. The estimated time-varying effects for FQA.

approach works well for practical situations. Also it was applied to an AD study to help identifying some time-varying risk factors for the AD conversion.

In the proposed procedure, two penalty functions, combined and separate penalty functions, were provided. For a practical problem, as discussed in Section 5, we suggest to first apply the latter on all coefficients to separate time-

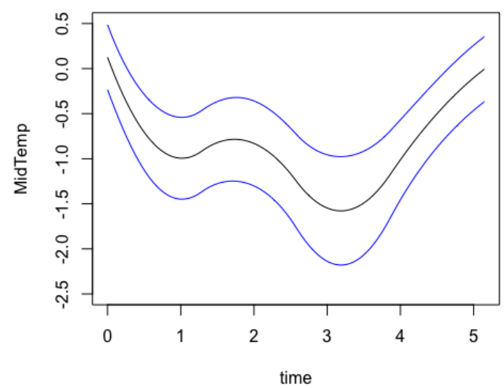


Figure 5. The estimated time-varying effects for MidTem.

varying coefficients from time-independent coefficients. The combined penalty function is then employed to provide more accurate estimation results. Also in the proposed approach, for simplicity, we only considered time-independent covariates and treated the baseline cumulative hazard function $\Lambda_0(t)$ as a step function. It is straightforward to generalize the presented method to the situation with time-dependent covariates but the implementation would be more complex. On estimation of $\Lambda_0(t)$, instead of treating it as a step function, one could also employ B-spline functions to approximate it as with time-varying coefficients. The method presented above will still be valid with some minor modifications.

ACKNOWLEDGEMENTS

The authors thank the editor and reviewers for helpful comments. We are also grateful to the ADNI community including the ADNI volunteers and their families.

Received 4 April 2021

REFERENCES

[1] Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand. J. Statist.* **30** 93–111. MR1963895

[2] Chen, M., Tong, X. W. and Sun, J. (2007). The proportional odds model for multivariate interval-censored failure time data. *Stat.Med* **26** 5147–5161. MR2412463

[3] Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **34(2)** 187–220. MR0341758

[4] Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics* **30** 74–99. MR1892656

[5] Han, X., Zhang, Y., Shao, Y. and Initiative., A. D. N. (2017). Application of concordance probability estimate to predict conversion from mild cognitive impairment to Alzheimer's disease. *Biostat Epidemiol.* **1** 105–118.

[6] Honda, T. and Härdle, W. K. (2014). Variable selection in Cox regression models with varying coefficients. *Journal of Statistical Planning and Inference* **148** 67–81. MR3174148

[7] Hu, H., Li, H., Li, J., Yu, J., Tan, L. and Initiative, A. D. N. (2018). Genome-wide association study identified ATP6V1H locus influencing cerebrospinal fluid BACE activity. *BMC Med Genet.* **19(75)** 1–8.

[8] Huang, J., Horowitz, J. L. and Wei, F. (2010). Variable selection in nonparametric additive models. *Annals of Statistics.* **38** 2282–2313. MR2676890

[9] Li, C., Pak, D. and Todem, D. (2020). Adaptive lasso for the Cox regression with interval censored and possibly left truncated data. *Statistical method in Medical Research* **29(4)** 1243–1255. MR4097142

[10] Li, K., Chan, W., Doody, R. S., Quinn, J., Luo, S. and Initiative, A. D. N. (2017). Alzheimers Disease Neuroimaging Initiative. Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to- event data. *J Alzheimers Dis.* **58(2)** 361–371.

[11] Murphy, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Statist.* **20** 35–50. MR1221960

[12] Rhinn, H., Fujita, R., Qiang, L., Cheng, R., Lee, J. H. and Abeliovich, A. (2013). Integrative genomics identifies APOE $\varepsilon 4$ effectors in Alzheimer's disease. *Nature.* **500** 45–50.

[13] Scolas, S., El Ghouch, a. L. C. A. and Oulhaj, A. (2016). Variable selection in a flexible parametric mixture cure model with interval-censored data. *Statistics in Medicine* **35** 1210–1225. MR3476537

[14] Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.*, First ed. Springer-Verlag New York. MR2287318

[15] Tian, L., Zucker, D. and Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association* **469** 172–183. MR2156827

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

[17] Wang, L., McMahan, C. S., Hudgens, M. G. and Qureshi, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics* **72** 222–231. MR3500591

[18] Wu, Q., Zhao, H., Zhu, L. and Sun, J. (2020). Variable selection for high-dimensional partly linear additive Cox model with application to Alzheimer's disease. *Statistics in Medicine* **39(23)** 3120–3134. MR4151923

[19] Wu, Y. and Cook, R. (2015). Penalized regression for interval-censored times of disease progression: selection of HLA markers in psoriatic arthritis. *Biometrics* **71** 781–791. MR3402614

[20] Yan, J. and Huang, J. (2012). Model selection for Cox Models with Time-Varying Coefficients. *Biometrics* **68(2)** 419–428. MR2959608

[21] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society,Series B: Statistical Methodology* **68** 49–67. MR2212574

[22] Zhang, H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association.* **106:495** 1099–1112. MR2894767

[23] Zhang, H. and Lu, W. (2007). Adaptive lasso for Cox's proportional hazards model. *Biometrika* **94** 691–703. MR2410017

[24] Zhao, H., Wu, Q., Li, G. and Sun, J. (2020). Simultaneous estimation and variable selection for interval-censored data with broken adaptive ridge regression. *Journal of the American Statistical Association* **115(529)** 204–216. MR4078457

Kaiyi Chen
Department of Statistics
University of Missouri
146 Middlebush Hall
Columbia, Missouri
USA 65211
E-mail address: kcvpc@mail.missouri.edu

Jianguo Sun
Department of Statistics
University of Missouri
146 Middlebush Hall
Columbia, Missouri
USA 65211
E-mail address: sunj@missouri.edu