Published in final edited form as:

Biometrics. 2012 June ; 68(2): 419–428. doi:10.1111/j.1541-0420.2011.01692.x.

Model Selection for Cox Models with Time-Varying Coefficients

Jun Yan^{1,2,3,*} and Jian Huang^{4,5}

¹Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

²Institute for Public Health Research, University of Connecticut Health Center, East Hartford, CT 06108, USA

³Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT 06269, USA

⁴Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA

⁵Department of Biostatistics, School of Public Health, University of Iowa, Iowa City, IA 52242, USA

Summary

Cox models with time-varying coefficients offer great flexibility in capturing the temporal dynamics of covariate effects on right censored failure times. Since not all covariate coefficients are time-varying, model selection for such models presents an additional challenge, which is to distinguish covariates with time-varying coefficient from those with time-independent coefficient. We propose an adaptive group lasso method that not only selects important variables but also selects between time-independent and time-varying specifications of their presence in the model. Each covariate effect is partitioned into a time-independent part and a time-varying part, the latter of which is characterized by a group of coefficients of basis splines without intercept. Model selection and estimation are carried out through a fast, iterative group shooting algorithm. Our approach is shown to have good properties in a simulation study that mimics realistic situations with up to 20 variables. A real example illustrates the utility of the method.

Keywords

B-spline; Group lasso; Varying-coefficient

1. Introduction

Cox models with time-varying coefficients offer great flexibility in assessing the temporal dynamics of covariate effects on right censored failure times. When a large number of covariates are available, it is important to select a subset of significant variables and the forms of their effect, time-varying or time independent. Therefore, an ideal model selection procedure for Cox models with time-varying coefficients should distinguish three kinds of covariates: 1) those not in the model; 2) those in the model with time-independent coefficients; and 3) those in the model with time-varying coefficients.

*jyan@stat.uconn.edu.

Supplementary Materials

Web Appendix, Tables, and Figures referenced in Sections 4 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

For standard Cox models with time-independent coefficients, effective variable selection techniques have been available. The Lasso approach, widely used in variable selection for linear regression models (Tibshirani, 1996), has been extended to Cox models (Tibshirani, 1997). Zhang and Lu (2007) further proposed adaptive Lasso where the penalty on each coefficient is weighted by the inverse magnitude of an initial estimate of the coefficient. Fan and Li (2002) proposed a general nonconcave penalized partial likelihood approach, extending their methods for linear models (Fan and Li, 2001). Both approaches have the oracle property; that is, the asymptotic distribution of an estimated coefficient is the same as that when it is known a priori which variables are in the model.

For Cox models with time-varying coefficients, model selection has not been extensively studied. In fact, the literature on model selection for varying coefficient in general appears to be limited. In the framework of smoothing spline analysis of variance, Lin and Zhang (2006) proposed a component selection and smoothing operator (COSSO) that replaces the squared norm penalty in traditional smoothing spline methods with L_1 norm. This approach was later extended to varying coefficient Cox models (Leng and Zhang, 2006). Li and Liang (2008) proposed a two-part variable selection approach for semiparametric regression models. Variables in the parametric component are selected with the method of Fan and Li (2001), with unknown nonparametric coefficients replaced with estimates obtained from maximizing kernel based local likelihood. Variables in the nonparametric component are selected with backward elimination via a sequence of generalized quasi-likelihood ratio test (Fan et al., 2001). For varying-coefficient models of repeated measurements, Wang et al. (2008) proposed a regularized estimation using smoothly clipped absolute deviation (SCAD) with nonparametric coefficients expanded by basis functions. For nonparametric additive models, Huang et al. (2010) approximated additive components with B-spline expansions and selected nonzero components by selecting the groups of coefficients in the expansion via adaptive group lasso. An interesting work that selects between constant coefficient and varying coefficient is Leng (2009), where the COSSO penalty was redesigned to distinguish the two types of coefficients; this approach, however, does not select between nonzero and zero coefficients. Most recently, Zhang et al. (2011) proposed an automatic approach to discover whether a covariate effect is linear or nonlinear in addition to whether it is nonzero with different penalty terms on linear and nonlinear effects. In the context of Cox models with time-varying coefficients, simultaneous selection between varying coefficient and fixed coefficient in addition to selection between nonzero and zero coefficient has not been studied.

Two main classes of approaches for varying coefficient Cox models have been studied in the literature. The penalized partial likelihood approach uses smooth functions for coefficients, maximizing the log partial likelihood with a penalty on the roughness of the coefficients (Zucker and Karr, 1990). The kernel-weighted partial likelihood approach finds point estimator at each time by maximizing a weighted “local” log partial likelihood function (Cai and Sun, 2003; Tian et al., 2005). We focus on the first class of models, where each time-varying coefficient is expanded over a B-spline basis. Each coefficient is then characterized by a set of basis coefficients which is further treated as two groups. The first group captures the time-independent, overall level of the covariate effect, while the second group captures the temporal changes relative to the overall level over time. We propose to select significant variables and the temporal dynamics of their effects by applying the group lasso approach (Yuan and Lin, 2006) over these groups of coefficients.

The rest of the article is organized as follows. An adaptive group lasso method with penalized partial likelihood based on B-splines is proposed in Section 2. Computation details of the proposed model selection procedure is presented in Section 3. Numerical studies on

finite sample performance of the procedure are summarized in Section 4. The method is applied to a real data in Section 5. A discussion concludes in Section 6.

2. Adaptive Group Lasso with B-Splines

Consider a random sample of size n . Let T_i^* be the failure time and C_i the censoring time of subject i , $i = 1, \dots, n$. Let $X_i = (X_{i1}, \dots, X_{ip})^T$ be the vector of covariate for subject i . Define $T_i = \min(T_i^*, C_i)$ and $\Delta_i = I(T_i^* \leq C_i)$. Assume that T_i^* and C_i are conditionally independent given X_i , and that the censoring scheme is non-informative. The observed data are independent and identically distributed copies $\{T_i, \Delta_i, X_i\}$, $i = 1, \dots, n$.

The Cox model with time-varying coefficients is

$$h(t|X_i) = h_0(t) \exp[X_i^T \beta(t)], \quad (1)$$

where h_0 is an unspecified baseline function, and $\beta(t)$ is $p \times 1$ vector of time-varying coefficients. Let $B_j(t)$'s, $j = 1, \dots, q-1$, $q > 1$, be a set of B-spline basis of $q-1$ degrees of freedom without intercept on a predetermined time interval $[0, \tau]$. Assume that $\beta(t)$ is expanded by the B-spline basis, $\beta(t) = \Theta R(t)$, where $R(t) = \{1, B_1(t), \dots, B_{q-1}(t)\}^T$ and Θ is a $p \times q$ matrix of parameters to be estimated. Therefore, each time varying coefficient $\beta_j(t) = \Theta_j R(t)$, $j = 1, \dots, p$, is determined by Θ_j , the j th row of parameter matrix Θ .

We decompose each $\beta_j(t)$ into two parts by partitioning Θ_j into two parts, each corresponding to a partition of $R(t)$. That is, we write $\Theta_j = (\Theta_{j,1}, \Theta_{j,-1})$, where $\Theta_{j,1}$ is the coefficient of the first component, one, in $R(t)$, and $\Theta_{j,-1}$ consists of the coefficients of the remaining components in $R(t)$, $\{B_1(t), \dots, B_{q-1}(t)\}$. The intercept $\Theta_{j,1}$ represents a time-independent, overall effect while $\Theta_{j,-1}$ determines the temporal changes in $\beta_j(t)$ relative to the intercept. Because of this construction, the B-spline basis $\{B_1(t), \dots, B_{q-1}(t)\}$ cannot contain any intercept. With package splines from base R (R Development Core Team, 2011), this can be obtained from function `bs` with `intercept = FALSE`. In our simulation and analysis, we used function `bs` with quadratic splines (degree = 2) with $q-1$ degrees of freedom (df = $q-1$), with equally spaced interior knots.

Let $\theta = \text{vech}(\Theta)$, the vectorization of Θ by row. Assuming no ties in the observed failure times, the log partial likelihood function is

$$l_n(\theta) = \sum_{i=1}^n \Delta_i \left[X_i^T \Theta F(T_i) - \log \left\{ \sum_{j \in R_i} \exp(X_j^T \Theta F(T_i)) \right\} \right], \quad (2)$$

where $R_i = \{k: T_k \leq T_i\}$ is the risk set at time T_i . We propose to estimate θ by minimizing the negative penalized log partial likelihood

$$Q_{\lambda_n}(\theta) = -l_n(\theta) + P(\theta; \lambda_n), \quad (3)$$

where $P(\theta; \lambda_n)$ is a penalty function that penalizes coefficient estimates in groups with a tuning penalty parameter λ_n (Yuan and Lin, 2006).

Suppose we partition θ into g groups, $\theta_1, \dots, \theta_g$. The penalty function is

$$P(\theta; \lambda_n) = \lambda_n \sum_{i=1}^g W_i \|\theta_i\|, \text{ where } W_i \text{ is a penalty weight for group } i. \text{ The weight } W_i \text{ can have}$$

group size p_i built-in as in (Yuan and Lin, 2006) and can be chosen adaptively as in Zhang and Lu (2007). In particular, we use

$$W_i = \sqrt{p_i} / \|\tilde{\theta}_i\|, \quad (4)$$

where p_i is the size of group i , and $\tilde{\theta}_i$ is some initial, consistent estimator of θ_i . This weight penalizes more if the groups size p_i is larger or if the norm $\|\tilde{\theta}_i\|$ is smaller.

To select significant variables and the temporal nature of their effects, we consider two ways to partition θ . The first way puts each row in Θ into a single group, which leads to penalty function

$$P(\theta; \lambda_n) = \lambda_n \sum_{j=1}^p W_j \|\Theta_j\|. \quad (5)$$

This penalty treats Θ_j as a whole group without distinguishing whether β_j can be described by a time-independent effect. We call the penalty in (5) combined penalty because each covariate coefficient is penalized via a single penalty. Significant variable can be selected but all selected variables are bound to have time-varying coefficients. The second way further separates each Θ_j into two groups, a time-independent part $\Theta_{j,1}$ and a time-varying part $\Theta_{j,-1}$, $j = 1, \dots, p$. The penalty function is

$$P(\theta; \lambda_n) = \lambda_n \sum_{j=1}^p \{W_{j1} \|\Theta_{j,1}\| + W_{j2} \|\Theta_{j,-1}\|\}, \quad (6)$$

where W_{j1} and W_{j2} are weights as in (4) computed with the new partition of Θ_j . This penalty is expected to pick up the difference between time-varying coefficient and time-independent coefficient, if a covariate coefficient is selected to be nonzero. We call the penalty in (6) separate penalty because the overall level and the temporal changes of each covariate coefficient are penalized separately. When $\Theta_{j,-1}$ is zero and $\Theta_{j,1}$ is nonzero, $\beta_j(t)$ is time-independent. When both $\Theta_{j,-1}$ and $\Theta_{j,1}$ are nonzero, $\beta_j(t)$ is time-varying. It is possible that $\Theta_{j,1}$ is zero while $\Theta_{j,-1}$ is nonzero, in which case, the coefficient $\beta_j(t)$ crosses zero.

Our model selection procedure is summarized as follows.

1. Minimize (3) with combined penalty (5) and weight $W_j = \sqrt{q}$, $j = 1, \dots, p$, to obtain $\hat{\theta}$.
2. Minimize (3) with combined penalty (5) and weight W_j , $j = 1, \dots, p$, computed from (4).
3. Minimize (3) with separate penalty (6) and weight W_{j1} and W_{j2} , $j = 1, \dots, p$, computed from (4).

The last step accomplishes the task to select significant variables and select the temporal nature of their effects at the same time. The second step is unnecessary, merely listed here for comparing results from combined penalty with those from separate penalty.

3. Computation

3.1 Iterative Group Shooting Algorithm

We propose an iterative group shooting algorithm to minimize $Q_{\lambda_n}(\boldsymbol{\theta})$ in (3). For a fixed penalty parameter λ_n and fixed weight $W_j, j = 1, \dots, g$, the algorithm is an adaptation of the iterative reweighted least squares (IRLS) procedure (Tibshirani, 1997; Zhang and Lu, 2007) to group penalty. Let $G = -\Delta I_n(\boldsymbol{\theta}) = -I_n(\boldsymbol{\theta})/\boldsymbol{\theta}$ and $H = -\Delta^2 I_n(\boldsymbol{\theta}) = -2I_n(\boldsymbol{\theta})/\boldsymbol{\theta}\boldsymbol{\theta}^\top$. Let $\mathbb{X}^\top$ be the Cholesky decomposition of H . Define pseudo response vector $\mathbb{Y} = (\mathbb{X}^\top)^{-1}\{H\boldsymbol{\theta} - G\}$. Then, a quadratic approximation of $Q_{\lambda}(\boldsymbol{\theta})$ is

$$\frac{1}{2}(\mathbb{Y} - \mathbb{X}\boldsymbol{\theta})^\top(\mathbb{Y} - \mathbb{X}\boldsymbol{\theta}) + \lambda_n \sum_{j=1}^g W_j \|\boldsymbol{\theta}_j\|. \quad (7)$$

This is a penalized least square problem. A necessary and sufficient condition for $\boldsymbol{\theta}$ to be a solution to the penalized least square (7) is (Yuan and Lin, 2006)

$$-\mathbb{X}_j^\top(\mathbb{Y} - \mathbb{X}\boldsymbol{\theta}) + \frac{\lambda_j \boldsymbol{\theta}_j}{\|\boldsymbol{\theta}_j\|} = 0, \quad \boldsymbol{\theta}_j \neq 0, \quad (8)$$

$$\|-\mathbb{X}_j^\top(\mathbb{Y} - \mathbb{X}\boldsymbol{\theta})\| \leq \lambda_j, \quad \boldsymbol{\theta}_j = 0, \quad (9)$$

where $\lambda_j = \lambda W_j$. The closed-form solution of Yuan and Lin (2006) is not applicable because $\mathbb{X}$ is not group orthonormal; $\mathbb{X}$ is a triangular matrix from a Cholesky decomposition.

The condition (8) is equivalent to

$$S_j = \left(\mathbb{X}_j^\top \mathbb{X}_j + \frac{\lambda_j}{\|\boldsymbol{\theta}_j\|} I_{p_j} \right) \boldsymbol{\theta}_j, \quad (10)$$

where $S_j = \mathbb{X}_j^\top(\mathbb{Y} - \mathbb{X}\boldsymbol{\theta}_{-j})$, with $\boldsymbol{\theta}_{-j} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_{j-1}^\top, \mathbf{0}^\top, \boldsymbol{\theta}_{j+1}^\top, \dots, \boldsymbol{\theta}_g^\top)^\top$. Consider the iteration

$$\boldsymbol{\theta}_j^{(1)} = \left(\mathbb{X}_j^\top \mathbb{X}_j + \frac{\lambda_j}{\|\boldsymbol{\theta}_j^{(0)}\|} I_{p_j} \right)^{-1} S_j. \quad (11)$$

This iteration is similar to the unified algorithm of Fan and Li (2002), except it is done for each group as in the shooting algorithm of Fu (1998). When indeed we have $\mathbb{X}_j^\top \mathbb{X}_j = I_{p_j}$, it reduces to the closed-form solution in Yuan and Lin (2006).

Our iterative shooting algorithm is summarized as follows.

1. Initialize with $\boldsymbol{\theta}^{(0)}$.
2. For each $j = 1, \dots, g$, obtain $\boldsymbol{\theta}_j^{(1)}$ from

$$\theta_j^{(1)} = \begin{cases} \left(\mathbf{X}_j^T \mathbf{X}_j + \frac{\lambda_j}{\|\theta_j^{(0)}\|} I_{p_j} \right)^{-1} S_j, & \|S_j\| > \lambda, \\ 0 & \|S_j\| \leq \lambda. \end{cases}$$

3. Let $\theta_j^{(0)} = \theta_j^{(1)}$ and repeat until convergence.

Note that when updating θ_j , S_j is computed with the most recent version of θ_{-j} .

This algorithm does not have the drawback that once a coefficient is shrunk to zero, it will stay at zero (Fan and Li, 2002, p.1354) because in each iteration, each coefficient is checked to see if it is nonzero based on the most recent estimate of θ_{-j} . Also, since this algorithm can be considered a special case of the block coordinate descent method, it is guaranteed to converge to a local minimizer (Tseng, 2001; Tseng and Yun, 2009). Because the negative log-partial likelihood and the penalty function are convex, the algorithm converges to a global minimizer. In our simulation studies, the algorithm usually converges in a few steps with starting values obtained from the last λ value under a moderate tolerance.

3.2 Choosing the Tuning Parameter

The tuning penalty parameter λ_n is estimated by generalized cross validation (GCV) (Craven and Wahba, 1979). We illustrate with the combined penalty function (5). The minimizer of (7) can be approximated by a ridge solution $(H + \lambda_n D)^{-1} \mathbf{X}^T \mathbf{Y}$, where matrix $D = \text{diag}\{\text{diag}(W_1/\|\theta_1\|), \dots, \text{diag}(W_g/\|\theta_g\|)\}$. Then, the number of effective parameters is approximated by $p(\lambda_n) = \text{tr}\{(H + \lambda_n D)^{-1} H\}$, and the GCV function is approximated by

$$\text{GCV}(\lambda_n) = \frac{-l_n(\theta)}{n\{1 - p(\lambda_n/n)\}^2}.$$

The optimal λ_n is chosen as the minimizer of GCV over a grid of λ_n values.

The flexibility of the B-spline basis is determined by its degrees of freedom q , which is in turn determined by the number and locations of interior knots. In our implementation, we used quadratic B-splines with interior knots equally spaced or placed on the sample quantiles of the observed failure times.

3.3 Likelihood Derivatives Evaluation

To minimize (3) using the iterative shooting algorithm, efficient evaluation of the derivatives of the log partial likelihood function (2) is needed. A naive way using standard software for time dependent covariate is to construct $p \times q$ pseudo time dependent covariates $X_j \otimes R(t)$, where $\otimes$ is the Kronecker product. This is computationally expensive for even moderate sample sizes because the time-dependent covariates needs to be constructed for each observed event time.

Taking advantage of Kronecker product, a fast routine suggested by Perperoglou et al.

(2006) can be used. The gradient of (2) is $\nabla l_n(\theta) = \sum_{i=1}^n \Delta_i(X_i - \bar{X}_i(\theta)) \otimes F(T_i)$, where

$$\bar{X}_i(\Theta) = \frac{\sum_{j \in R_i} X_j \exp(X_j^\top \Theta F(T_i))}{\sum_{j \in R_i} \exp(X_j^\top \Theta F(T_i))}$$

is the mean of covariate X_j in risk set R_i weighted by $\exp(X_j^\top \Theta F(T_i))$. The Hessian matrix is $\nabla^2 l_n(\theta) = \sum_{i=1}^n \Delta_i C_i(\Theta) \otimes \{F(T_i) F^\top(T_i)\}$, where $C_i(\Theta)$ is the covariance matrix of covariate vector X_j in risk set R_i weighted again by $\exp(X_j^\top \Theta F(T_i))$. As shown by Perperoglou et al. (2006) and the numerical study in this article, such formulation is very efficient for larger sample sizes.

3.4 Variance Estimation

Following Fan and Li (2002), when the algorithm converges, the estimator satisfies the iteration

$$\widehat{\theta}^{(1)} = \widehat{\theta}^{(0)} - \left\{ \nabla^2 l_n(\widehat{\theta}^{(0)}) + \sum (\widehat{\theta}^{(0)}; \lambda_n) \right\}^{-1} \left\{ \nabla l_n(\widehat{\theta}^{(0)}) + U(\widehat{\theta}^{(0)}, \lambda_n) \right\},$$

where

$$\sum(\theta; \lambda_n) = \text{diag} \left\{ \frac{\lambda_n W_1}{\|\theta_1\|} I(p_1), \dots, \frac{\lambda_n W_g}{\|\theta_g\|} I(p_g) \right\},$$

$I(k)$ is identity matrix of dimension k ,

$$U(\theta; \lambda_n) = \text{diag} \left\{ \frac{\lambda_n W_1}{\|\theta_1\|} \text{diag}(\theta_1), \dots, \frac{\lambda_n W_g}{\|\theta_g\|} \text{diag}(\theta_g) \right\}.$$

The corresponding sandwich formula can be used as an estimator for the covariance of the estimate $\widehat{\theta}_{NZ}$, the nonzero component of $\widehat{\theta}$. That is, $\widehat{\text{cov}}(\widehat{\theta}_{NZ}) = A^{-1} B A^{-1}$, where

$$A = \nabla^2 l_n(\widehat{\theta}_{NZ}, 0) + \sum (\widehat{\theta}_{NZ}; \lambda_n)$$

and $B = \widehat{\text{cov}} \left[\nabla l_n(\widehat{\theta}_{NZ}, 0) \right]$.

Once the variance estimator of $\widehat{\theta}$ is obtained, the variance estimator of a nonzero coefficient $\beta_j(t)$ is then $F^\top(t) \widehat{\text{cov}}(\widehat{\theta}_j) F(t)$, where $\widehat{\text{cov}}(\widehat{\theta}_j)$ is the variance estimator of θ_j extracted from the estimated variance matrix of $\widehat{\theta}$. This estimator can be used to construct pointwise confidence intervals for $\beta_j(t)$.

4. Numerical Studies

Simulations were conducted to study the performance of the proposed adaptive group lasso with B-splines for finite samples. In particular, we want to check if the proposed method can 1) pick up important variables correctly (in the model or not); and 2) pick up the form of important variables correctly (time-varying versus time-independent).

Four factors are considered in our simulation design: number of covariates (10 and 20), censoring percentage c_p (20% and 40%), sample size n (200 and 400), and effect scale s (1 and 2). The effect scale — a multiplier on all the coefficients — is designed to study the influence of effect size or signal level on the performance of the proposed methods.

Event times are generated from a varying-coefficient Cox model (1) with time-independent covariate vector X and coefficients $\beta(t)$, whose nonzero components are $\beta_2(t) = -s\{1 + \cos(\pi t)\}I(0 < t < 1)$, $\beta_3(t) = s\{0.5 + \sin(\pi t/2)\}$, and $\beta_8(t) = -s$; see Figure 1 for $t \in (0, 2)$. That is, out of 10 or 20 covariates, the 2nd and 3rd ones have time-varying coefficients, the 8th one has time-independent coefficient, and all the rest have coefficient zero. Note that $\beta_2(t)$ diminishes to zero at $t = 1$ and remains zero afterwards, which makes model selection and estimation harder. The baseline hazard function also has effect scale s built in, $\lambda(t) = \exp\{-s \cos(\pi t/2)\}$. Covariate vector X is generated from a multivariate normal distribution whose marginals are all $N(0, 0.5)$, and whose pairwise correlation coefficients are $0.5^{|j-k|}$ for pair (j, k) . Censoring times are generated from a mixture of uniform distribution over $(0, 2)$ and a point mass at 2, with the mixing probability calibrated to yield desired censoring percentage c_p . For each scenario, 100 datasets are generated.

Given a simulated dataset, we use quadratic B-splines with 5 degrees of freedom, with equally spaced knots in time window $(0, 2)$, for each covariate coefficient. This gives two equal distant interior knots in $(0, 2)$. Model selection results are obtained from combined penalty (5) and separate penalty (6), denoted as Method 1 and Method 2, respectively. For comparison, we also report the model selection results from adaptive lasso with all covariate coefficients specified as time-independent, which is denoted as Method 0.

Table 1 and Table 2 summarize the variable selection results for 10 covariates and 20 covariates, respectively, regardless of the time nature of their effects. We report the frequency of each variable selected, the average number of groups selected (NG), and the average MSE over 100 replicates. The “correct” NGs are 3, 3, and 5 for Method 0, Method 1, and Method 2, respectively. The MSE at a specific time t is calculated as $\{\hat{\beta}(t) - \beta(t)\}^T V \{\hat{\beta}(t) - \beta(t)\}$, where V is the population covariance matrix of the covariates. The reported MSE is the average of pointwise MSE over a equally spaced grid of 100 points in time interval $(0, 2)$.

In all scenarios, both Method 1 and Method 2 seem to work reasonably well in selecting covariates X_3 and X_8 . Covariate X_2 with a diminishing effect is difficult to select. It is selected more frequently with Method 2 than with Method 1, which is expected as the effect is easier to be picked up as time-independent with separate penalty. As the effect scale s increases from 1 to 2, both Method 1 and Method 2 selects more often X_2 , and the selection of non-important variables becomes less often or does not get worse. This is not true, however, for Method 0, which selects more often X_2 but at the same time, selects more often non-important variables. All methods appear to improve as the sample size increases. For sample size $n = 200$, Method 2 performs similar to Method 0 in that non-important variables are over-selected. As sample size increases, the advantage of Method 2 relative to Method 0 becomes evident with less over-selection and smaller MSEs. For instance, in the scenario of $s = 1$, $n = 400$, and $c_p = 40\%$ with 20 covariates, the MSE is 0.447 for Method 2 and 0.883 for Method 0; the MSE of Method 1 is 0.512, which is in between the other two. As

censoring gets heavier, correct selection of X_2 improves and the overall MSE decreases for both Method 1 and Method 2. This may be explained by the fact that when censoring is heavier, the proportion of events in earlier times is higher, which increases the chances of the earlier part of β_2 being selected into the model as a negative time-independent effect.

It is of particular interest to check if Method 2 can tell whether a coefficient is time-independent or not. Table 3 summarizes these results for the three variables with nonzero coefficient. We report the frequencies that the intercept (Int) component and the time-varying (TV) component of each nonzero effect is selected. The performance of Method 2 improves as the effect scale or the sample sizes increases, with much higher frequency that X_2 is selected to have a time-varying effect. Covariate X_3 , which has a positive bump effect, is selected to have a time-varying effect about 2/3 of the times or more for $s = 1$, and almost all the time for $s = 2$. Covariate X_8 is correctly selected to have a time-independent effect most of the time even at sample size $n = 200$. For instance, consider again the scenario of $s = 1$, $n = 400$, and $c_p = 40\%$ with 20 covariates. Variables X_2 and X_3 are correctly selected to have time-varying coefficient for 84 and 80 times, respectively; variable X_8 is incorrectly selected to have time-varying coefficient only 4 times.

Finally, to study recovery of the nonzero coefficients, we plot in Figure 1 the 100 estimated coefficient curves overlaid with the true curves for the scenarios with $n = 200$ and $c_p = 40\%$, using both combined penalty and separate penalty. It is clear that the scale size s plays an important role here. For stronger signal ($s = 2$), the estimated curves are much closer to and tighter around the true curves for all three coefficients. In particular, for $s = 2$, estimates of the diminishing effect $\beta_2(t)$ are recovering the true curve reasonably well; for $s = 1$, however, the earlier negative effect is more obviously shrunk to zero, and in the case of separate penalty, many of the estimates are estimated as negative but time-independent. With separate penalty, estimates of $\beta_3(t)$ are all time-varying in the case of $s = 2$, but with a noticeable number of time-independent curves in the case of $s = 1$. The separate penalty performs very well in estimating the time-independent coefficient $\beta_8(t)$, and gives less bias in comparison with those estimates from the combined penalty. By comparing the estimated curves under combined penalty and under separate penalty, it seems that when the true curve is time-independent, as with $\beta_8(t)$, the separate penalty gives lighter shrinkage towards zero and less variability; when the true curve is time-varying, however, as with $\beta_2(t)$ or $\beta_3(t)$, the combined penalty seems to provide less variability. This is observed in both cases of $s = 1$ and $s = 2$. The observation may be expected since the separate penalty approach tries to achieve more than the combined penalty, and it comes with a cost because, when the true curves are time-varying, there is a chance that the separate penalty may not select the necessary intercept as seen in Table 3.

Also plotted in Figure 1 are the averages of the 100 estimated coefficients curves and their pointwise 95% confidence intervals constructed using the variance estimator in Section 3.4. The standard errors appear to underestimate the true variation, which may be related to the shrinkage effect in the estimation. The underestimation of variation was also observed for Cox model with constant coefficients (Zhang and Lu, 2007). In our setting, the number of parameters in Θ is even more and, hence, an even larger sample size is necessary for the asymptotic variance to provide good approximation.

The performance of the methods is more aggressively studied by replacing the diminishing effect $\beta_2(t)$ with a crosszero effect $\beta_2(t) = -s \cos(\pi t/2)$, which makes the problem much harder since $\beta_2(t)$ integrates to zero over $(0, 2)$. Results in analogy to Tables 1–3 and Figure 1 are reported in the Web Appendix. In this study, the crosszero effect is very hard to be picked up with $s = 1$; for example, with $n = 400$, $c_p = 40\%$ and 20 covariates in Web Table A.2, only 19 and 38 out of 100 times X_2 is selected by Method 1 and Method 2,

respectively. With $s = 2$, these frequencies increase to 96 and 98, respectively, and further, it was selected as time-varying coefficient 94 and 98 times, respectively (Web Table A.3). The estimated $\beta_2(t)$ curves in Web Figure A.1 are not close to the true curve with $s = 1$. Nevertheless, with $s = 2$, the estimates are recovering the true curve reasonably well from both Method 1 and Method 2, albeit shrunken towards zero at the two endpoints where it was most away from zero. Observations about recovering $\beta_3(t)$ and $\beta_8(t)$ are similar to those seen in Figure 1. The poor performance of $\hat{\beta}_2(t)$ in the case of $s = 1$ is not a surprise because the problem is a much harder one. As the signal gets stronger, our methods can be useful in detecting and estimating such crosszero effects.

5. The Primary Biliary Cirrhosis Data

We apply the proposed method to the primary biliary cirrhosis (PBC) data, which has been analyzed in the context of model selection for Cox model with time-independent coefficients (Tibshirani, 1997; Zhang and Lu, 2007). PBC is a rare but fatal chronic autoimmune liver disease, with a prevalence of about 50-cases-per-million population (Fleming and Harrington, 1991). The dataset contains followup of 312 randomized and 106 unrandomized patients with PBC at Mayo Clinic between January 1974 and May 1984. The dependence of survival time on 17 covariates is studied in a Cox model with possibly time-varying coefficients. The survival time is the number of days between registration and the earlier of death or study analysis time in 1986. We consider the 312 randomized patients and, after removing missing values, end up with 276 observations. The 17 covariates are, in the same order as in Tibshirani (1997), 1) trt, treatment indicator (1 = treatment); 2) age (in 10 years); 3) female, gender indicator (1 = female); 4) ascites, presence of ascites; 5) hypato, presence of hypatomegaly; 6) spiders, presence of spiders; 7) edema, severity of oedema; 8) logbili, logarithm of serum bilirubin (mg/dl); 9) chol, serum cholesterol (mg/dl); 10) logalb, logarithm of albumin (g/dl); 11) copper, urine copper (mg/day); 12) alk.phos, alkaline phosphatase (U/l); 13) ast, aspartate aminotransferase (U/ml); 14) trig, triglycerides (mg/dl); 15) platelet, platelet count per 10^{-3} ml³; 16) logprotime, logarithm of prothrombin time (sec); 17) stage, histologic stage of disease (graded 1, 2, 3, or 4). Note that, we took log on serum bilirubin, albumin, and prothrombin time since Tian et al. (2005) and Martinussen and Scheike (2002) found possibly time-varying coefficients for these covariates.

We first fit the Cox model with time-independent coefficients for all 17 covariates without any penalty. The inverse of the absolute value of these estimates were then used as weights in two adaptive procedures, adaptive lasso (ALASSO) with time-independent coefficients and the proposed adaptive group lasso (AGLASSO) with B-splines. The ALASSO approach is the same as that in Zhang and Lu (2007), except that we took log for three aforementioned covariates. The AGLASSO approach allows time-varying coefficients and, for each covariate, penalizes the time-independent part and the time-varying part of the coefficient in separate groups. The B-spline basis are quadratic with 5 degrees of freedom over the time interval of (0, 3200) days, where 3200 is approximately the 90th percentile of the observed event times. After a final model is selected from ALASSO or AGLASSO, we then fit a Cox model without any penalty assuming that the selected models are known. Table 6 summarizes the results. Both ALASSO and AGLASSO selected the same set of covariates. The only variable selected as having time-varying coefficient was logbili, which is consistent with the finding of Tian et al. (2005). The estimate and pointwise 95% confidence interval of this coefficient is plotted in Figure 2. The estimated effect has a bump during between days 1000 and 1500, after which it diminishes gradually. Two other covariates, logprotime and edema, were identified as possibly having time-varying coefficient by Tian et al. (2005), who used only 5 out of the 17 variables to start with. Using AGLASSO, these two variables were selected to be significant in the model but their effects were not found to be time-varying.

6. Discussion

Variable selection for semiparametric model is different from traditional variable selection for linear models in that the temporal nature of the coefficient of each selected variable needs to be selected as well. The method of Li and Liang (2008), developed for generalized varying-coefficient partially linear models, can be extended to varying coefficient Cox models, in which case, nonparametric coefficients would be fitted with kernel-based local partial likelihood. Nevertheless, this method assumes knowledge a priori about which covariates have varying-coefficient. Our nonparametric coefficients are fitted with smooth functions expanded using B-spline basis. Penalizing a time-independent part and a time-varying part separately for each coefficient, our adaptive group lasso approach not only selects significant variables but also identifies which ones have varying-coefficient. In addition to selecting those variables that are important. This is important for practitioners who do not have prior knowledge or are not willing to make assumptions about the functional form of the covariate coefficients. Our simulation studies shows rather good results for sample size as big as 400 with moderate censoring in selecting 3 important variables out of 20. The working version of our implementation as an informal R package is available upon request.

Our focus is on the methodology development, its computational implementation, and numerical evaluation of its performance. An important question that we have not addressed is the estimation and selection consistency of the proposed method. This is an interesting and challenging problem, especially if p is allowed to diverge with n . We conjecture that the procedure can correctly distinguish time-varying and time-independent covariates correctly as the sample size goes to infinity, in light of the results of (Huang et al., 2010) for nonparametric additive models. A rigorous proof, however, is not straightforward. The main difficulty arises from the fact that the log-partial likelihood is not a sum of independent terms. Therefore, the tools (e.g., maximal inequalities for independent random variables) from the empirical process theory are not applicable. The martingale method that is effective in studying Cox models with time-independent covariates does not apply to the current problem either. Research to carefully address all the technical details is warranted.

The proposed methods raise several questions. A multiple degree of freedom factor would lead to a collection of groups, each one formed by splines basis corresponding to one degree of freedom. For instance, the histologic stage of disease in our analysis of the PBC data was treated as a numerical variable, but it could, even preferably sometimes, be treated as a factor. A naive solution would be to treat all the groups as one big group and then apply the proposed method; this way, all contrasts of this factor are either in or out of the model altogether. A better solution would be to add different penalties to different levels of grouping similar to the bi-level penalty of (Breheny and Huang, 2009). It is known that a nonlinear effect in a Cox model may be mis-identified as time-varying effect (Therneau and Grambsch, 2000). Model selection with nonlinear effects may be done with the fractional polynomials approach (Royston and Altman, 1994; Sauerbrei and Royston, 1999). A sensitivity study of the performance of the proposed method under Cox models with nonlinear effect would be interesting. Comparison with nonconcave penalty approaches such as group SCAD (Fan and Li, 2001) and group minimax concave penalty (MCP) (Zhang, 2010) is of great interest as always. Our computing algorithm, however, is built upon the Karush–Kuhn–Tucker conditions of group LASSO (Yuan and Lin, 2006), which makes it nontrivial to adapt to SCAD and MCP. The coordinate descent algorithm for group SCAD and group MCP in (Breheny and Huang, 2011) may be extended to handle groups of basis coefficients and to the context of Cox models with varying-coefficients. Such extensions, implementation, and their numerical performance, however, deserve separate manuscripts on themselves.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Yan's research was partially supported by U.S. National Science Foundation grant DMS 0805965. Huang's research was partially supported by NIH grants R01CA120988, R01CA142774 and NSF grant DMS 0805670. The computing was facilitated by a beowulf cluster at the Department of Statistics, University of Connecticut, acquired under the partial support of NSF SCREMS grant 0723557.

References

- Breheny P, Huang J. Penalized methods for bi-level variable selection. *Statistics and its interface*. 2009; 2:369–380. [PubMed: 20640242]
- Breheny P, Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*. 2011; 5:232–253. [PubMed: 22081779]
- Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*. 2003; 30:93–111.
- Craven P, Wahba G. Smoothing noisy data with spline functions. *Numerische Mathematik*. 1979; 31:377–403.
- 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.
- Fan J, Zhang C, Zhang J. Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*. 2001; 29:153–193.
- Fleming, TR.; Harrington, DP. *Counting Processes and Survival Analysis*. John Wiley & Sons; 1991.
- Fu WJ. Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*. 1998; 7:397–416.
- Huang J, Horowitz JL, Wei F. Variable selection in nonparametric additive models. *Annals of Statistics*. 2010; 38:2282–2313. [PubMed: 21127739]
- Leng C. A simple approach for varying-coefficient model selection. *Journal of Statistical Planning and Inference*. 2009; 139:2138–2146.
- Leng C, Zhang HH. Model selection in nonparametric hazard regression. *Journal of Nonparametric Statistics*. 2006; 18:417–429.
- Li R, Liang H. Variable Selection in semiparametric regression modeling. *Annals of Statistics*. 2008; 36:261–286. [PubMed: 19122808]
- Lin Y, Zhang HH. Component selection and smoothing in multivariate nonparametric regression. *Annals of Statistics*. 2006; 34:2272–2297.
- Martinussen T, Scheike TH. A flexible additive multiplicative hazard model. *Biometrika*. 2002; 89:283–298.
- Perperoglou A, le Cessie S, van Houwelingen HC. A fast routine for fitting Cox models with time varying effects of the covariates. *Computer Methods and Programs in Biomedicine*. 2006; 25:154–161. [PubMed: 16426701]
- R Development Core Team. *R: A Language and Environment for Statistical Computing*; Vienna, Austria. R Foundation for Statistical Computing; 2011.
- Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics*. 1994; 43:429–467.
- Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1999; 162:71–94.

- Therneau, TM.; Grambsch, PM. Modeling Survival Data: Extending the Cox Model. Springer-Verlag Inc; 2000.
- Tian L, Zucker D, Wei L. On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*. 2005; 100:172–183.
- Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B: Methodological*. 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]
- Tseng P. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*. 2001; 109:475–494.
- Tseng P, Yun S. Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization. *Journal of Optimization Theory and Applications*. 2009; 140:513–535.
- Wang L, Li H, Huang JZ. Variable selection in nonparametric vary-coefficient models for analysis of repeated measurements. *Journal of American Statistical Association*. 2008; 103:1556–1569.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*. 2006; 68:49–67.
- Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*. 2010; 38:894–942.
- Zhang HH, Cheng G, Liu Y. Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association*. 2011 Forthcoming.
- Zhang HH, Lu W. Adaptive lasso for Cox's proportional hazards model. *Biometrika*. 2007; 94:691–703.
- Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *The Annals of Statistics*. 1990; 18:329–353.

Figure 1.

Estimated curves (gray) of the three nonzero coefficients from 100 replicates in the scenario of sample size $n = 400$ and censoring percentage $c_p = 40\%$ with 20 covariates when $\beta_2(t) = \mathbb{I}(0 < t < 1)s\{1 - \cos(\pi t)\}$. The dark lines are the true curves. The dashed lines are the average of 100 estimates. The dotted lines are the pointwise 95% confidence intervals.

Figure 2.
Time-varying coefficient estimate of covariate logbili

Table 1

Model selection results from 100 runs with 10 covariates. The three entries in each table cell are the counts of each variable being selected in time-fixed-coefficient models (Method 0), combined-penalty-varying-coefficient models (Method 1), and separate-penalty-varying-coefficient models (Method 2), respectively.

n	c_p	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	NG	MSE
$s = 1$													
200	20	7	51	100	3	4	2	4	100	1	5	2.8	1.053
		1	32	99	1	1	0	1	100	1	0	2.4	1.014
		8	63	100	2	5	2	4	100	1	5	3.7	0.852
	40	11	67	96	5	8	9	11	98	13	6	3.2	1.123
400		4	63	96	0	0	2	3	97	5	1	2.7	0.937
		11	82	98	3	6	8	8	98	10	6	4.2	0.845
	20	9	87	100	2	6	1	1	100	3	0	3.1	0.870
		1	73	100	0	0	0	0	99	0	0	2.7	0.692
400		6	94	100	0	1	0	1	99	0	0	4.3	0.591
		10	95	100	6	2	4	1	100	4	7	3.3	0.894
		2	94	100	0	0	1	1	100	0	0	3.0	0.555
		10	97	100	3	1	2	2	100	4	5	4.7	0.503
$s = 2$													
200	20	6	98	100	8	9	7	8	100	7	4	3.5	3.421
		0	87	100	0	0	1	1	100	1	0	2.9	2.030
		2	98	100	3	2	3	2	100	3	0	4.8	1.631
	40	10	100	100	21	12	10	5	100	15	11	3.8	3.491
400		0	100	100	0	0	0	1	100	1	0	3.0	1.442
		4	100	100	9	2	2	3	100	8	4	5.1	1.311
	20	14	100	100	2	7	4	3	100	3	3	3.4	3.230
		0	100	100	0	0	1	0	100	0	0	3.0	0.800
400		1	100	100	1	2	2	0	100	1	1	5.0	0.780
	40	19	100	100	15	10	12	17	100	12	20	4.0	3.289
		0	100	100	0	0	0	0	100	0	0	3.0	0.774
		4	100	100	2	0	1	2	100	3	0	5.0	0.706

NG: number of selected groups. MSE: mean squared error.

Table 2

Model selection results from 100 runs with 20 covariates. The three entries in each table cell are the counts of each variable being selected in time-independent-coefficient models (Method 0), combined-penalty-varying-coefficient models (Method 1), and separate-penalty-varying-coefficient models (Method 2), respectively.

n	c_p	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	NG	MSE
$s=1$																							
200	20	8	51	100	6	3	3	6	100	3	2	6	3	4	5	8	6	3	4	1	4	3.3	1.046
		1	30	100	2	0	0	2	98	1	0	1	0	0	0	0	0	0	0	0	1	2.4	1.037
		7	56	100	6	5	1	5	100	3	1	7	2	6	2	5	3	1	2	3	4	3.9	0.889
	40	6	69	96	7	7	6	9	100	9	6	6	4	7	9	9	4	6	8	4	7	3.8	1.125
400	20	2	64	97	3	3	2	5	100	1	3	2	0	2	1	1	0	2	0	1	4	2.9	0.922
		7	80	97	6	8	4	5	100	9	6	6	3	4	9	5	3	3	6	5	9	4.8	0.856
		2	95	100	1	5	5	4	100	4	4	0	0	4	1	3	3	1	3	0	5	3.4	0.827
		0	82	100	0	0	0	1	100	0	0	0	0	0	0	0	0	0	0	0	0	2.8	0.614
		1	98	100	0	0	2	2	100	0	0	0	0	2	0	2	0	0	1	1	2	4.5	0.510
	40	8	97	100	3	4	3	6	100	5	5	3	6	8	5	6	4	9	5	6	7	3.9	0.883
		0	98	100	0	0	0	1	100	0	0	0	0	0	0	0	1	0	0	1	0	3.0	0.512
		5	100	100	1	3	4	3	100	1	4	1	4	7	3	4	3	6	3	5	4	5.2	0.447
$s=2$																							
200	20	11	99	100	9	4	6	13	100	8	2	5	10	8	8	7	6	4	6	3	4	4.1	3.362
		0	93	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	2.9	1.936
		1	98	100	2	0	0	1	100	0	1	0	2	2	0	1	0	0	0	0	0	4.6	1.680
	40	15	99	100	16	9	14	11	100	12	13	11	9	13	10	11	13	15	10	9	11	5.0	3.418
400		1	99	100	1	0	0	3	100	2	0	0	0	0	0	0	1	0	0	0	0	3.1	1.402
		7	100	100	5	3	3	7	100	5	3	4	4	4	2	4	3	4	0	5	2	5.4	1.243
		10	100	100	7	6	5	10	100	5	5	6	2	3	11	7	9	6	7	5	7	4.1	3.210
		0	100	100	0	0	0	1	100	0	0	0	0	0	0	0	0	0	0	0	0	3.0	0.812
		2	100	100	1	1	2	3	100	1	1	4	1	0	3	1	3	1	2	3	0	5.2	0.762
	40	18	100	100	10	14	17	20	100	5	15	19	13	18	8	16	14	12	13	9	16	5.4	3.259
		0	100	100	0	0	1	0	100	0	0	0	0	0	0	0	0	0	1	0	0	3.0	0.749
		7	100	100	4	2	3	6	100	0	2	5	5	2	3	5	6	1	2	2	3	5.6	0.682

NG: number of selected group. MSE: mean squared error.

Time-varying selection results of separate-penalty-varying-coefficient models (Method 2) for variables 2, 3, and 8 from 100 runs.

Table 3

n	c _p	10 covariates						20 covariates					
		X ₂		X ₃		X ₈		X ₂		X ₃		X ₈	
		Int	TV	Int	TV	Int	TV	Int	TV	Int	TV	Int	TV
s = 1													
200	20	63	12	97	67	100	1	56	11	96	64	100	0
	40	81	26	86	76	98	3	80	36	87	71	100	4
400	20	94	43	94	89	99	1	98	56	93	91	100	1
	40	97	71	89	82	100	1	100	84	90	80	100	4
s = 2													
200	20	98	79	85	98	100	1	98	78	75	100	100	0
	40	100	95	83	97	100	4	100	97	79	99	100	3
400	20	100	100	85	100	100	2	100	100	90	100	100	3
	40	100	100	86	100	100	3	100	100	88	100	100	10

Int: the intercept component. TV: the time-varying component.

Table 4

Estimated coefficients and standard errors from ML, ALASSO, AGLASSO for the PBC data. Results for ALASSO and AGLASSO were obtained from refitting the selected model without penalty.

Covariate	ML		ALASSO		AGLASSO	
	Coef	Std.Err	Coef	Std.Err	Coef	Std.Err
trt	-0.062	0.211				
age	0.261	0.113	0.270	0.124	0.263	0.126
female	-0.256	0.317				
ascites	0.162	0.381				
hepato	-0.100	0.254				
spiders	0.049	0.243				
edema	0.926	0.378	0.842	0.410	0.932	0.443
logbili	0.723	0.162	0.699	0.115	See Figure 2	
chol	0.000	0.000				
logalb	-2.270	0.947	-2.538	0.762	-2.440	0.789
copper	1.694	1.251	2.218	1.236	2.089	1.261
alk.phos	0.000	0.000				
ast	0.003	0.002				
trig	-0.002	0.001				
platelet	0.001	0.001				
logprotime	2.335	1.321	2.099	1.241	1.822	1.249
stage	0.381	0.176	0.274	0.140	0.278	0.143