DEPARTMENT OF STATISTICS

North Carolina State University

2501 Founders Drive, Campus Box 8203

Raleigh, NC 27695-8203

Institute of Statistics Mimeo Series No. 2579

Adaptive-LASSO for Cox's Proportional Hazards Model

Hao Helen Zhang

Wenbin Lu

Department of Statistics, North Carolina State University, Raleigh, NC

hzhang@stat.ncsu.edu, lu@stat.ncsu.edu

Supported in part by National Science Foundation grants DMS-0405913 and DMS-0504269.

Adaptive-LASSO for Cox's Proportional Hazards Model

By Hao Helen Zhang and Wenbin Lu

Department of Statistics, North Carolina State University, North Carolina 27695, U.S.A.

hzhang@stat.ncsu.edu, lu@stat.ncsu.edu

Summary

We investigate the variable selection problem for Cox's proportional hazards model, and propose a unified model selection and estimation procedure with desired theoretical properties and computational convenience. The new method is based on a penalized log partial likelihood with the adaptively-weighted L_1 penalty on regression coefficients, and is named adaptive-LASSO (ALASSO) estimator. Instead of applying the same penalty to all the coefficients as other shrinkage methods, the ALASSO advocates different penalties for different coefficients: unimportant variables receive larger penalties than important variables. In this way, important variables can be protectively preserved in the model selection process, while unimportant ones are shrunk more towards zero and thus more likely to be dropped from the model. We study the consistency and rate of convergence of the proposed estimator. Further, with proper choice of regularization parameters, we have shown that the ALASSO perform as well as the oracle procedure in variable selection; namely, it works as well as if the correct submodel were known. Another advantage of the ALASSO is its convex optimization form and convenience in implementation. Simulated and real examples show that the ALASSO estimator compares favorably with the LASSO.

Some key words: Adaptive LASSO (ALASSO), LASSO, Penalized partial likelihood, Proportional hazards model, Variable selection.

1 Introduction

One main issue in time-to-event data analysis is to study the dependence of survival time T on covariates $\mathbf{z} = (z_1, \cdots, z_d)$. Cox's proportional hazards model (Cox 1972, 1975) is one of the most popular models in literature and has been widely studied. To be specific, the hazard function $h(t|\mathbf{z})$ of a subject with covariates $\mathbf{z}$ is specified by

$$h(t|\mathbf{z}) = h_0(t) \exp(\sum_{j=1}^{d} z_j \beta_j), \tag{1}$$

where $h_0(t)$ is a completely unspecified baseline hazard function and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_d)'$ is an unknown vector of regression coefficients.

In practice, the number of covariates d is often large and not all the covariates contribute to the prediction of survival outcomes. Since some components of $\boldsymbol{\beta}$ may be zero, the regression relationship can be described by a more compact model. An effective variable selection often leads to parsimonious models with better risk assessment and easy interpretation. Two desired criteria for variable selection are:

(i) The selected model contains all nonzero components of $\boldsymbol{\beta}$;

(ii) The selected model has the smallest size among all the models satisfying (i).

Criterion (i) implies that no important variables are missed, while (ii) defines the optimality of the model. When the sample size goes to infinity, an ideal model selection and estimation procedure should be able to, on one side, identify the optimal model with probability one, and on the other side, provide consistent and efficient estimates for the coefficients of important variables. In this paper, we propose a new procedure, the adaptive LASSO estimator, and show that it satisfies all these theoretical properties.

Many variable selection techniques for linear regression models have been extended to the context of survival models. Two popular procedures are the best subset selection and stepwise selection. Another class of methods are asymptotic procedures based on score tests, Wald tests, and other approximate chi-square testing procedures. Bootstrap sampling procedures for variable selection were studied in Sauerbrei & Schumacher (1992). Bayesian variable selection for survival data was investigated by Faraggi & Simon (1998) and Ibrahim, Chen & MacEachern (1999). However, the theoretical properties of these methods are generally unknown (Fan & Li 2002).

Recently a family of penalized partial likelihood methods, such as LASSO (Tibshirani 1997) and SCAD (Fan & Li, 2002), have been proposed for Cox's proportional hazards model. These methods shrink some coefficients to exactly zeros, and hence simultaneously select important variables and estimate regression coefficients. The LASSO is simple and popular but does not have the oracle properties (Fan & Li, 2002). On the other hand, the SCAD has very nice theoretical properties but its penalty form is not convex, which makes the method computationally difficult. In this work, we have developed a new procedure, the adaptive LASSO (ALASSO), which combines the advantages of aforementioned methods and overcomes their drawbacks. Compared with the LASSO, the ALASSO solutions enjoy the theoretical properties like root-n consistency and oracle properties, resulting from the use of adaptive penalties. Compared with the SCAD, the ALASSO penalty has a convex formulation similar to the LASSO, and hence is much easier to solve in practice.

The ALASSO method is based on a penalized partial likelihood with the adaptively-weighted L_1 penalties on regression coefficients. Unlike the LASSO and SCAD methods which apply the same penalty to all the coefficients, the ALASSO allows the penalty on each coefficient to be individually adjusted so that unimportant covariates receive larger penalties than important ones. To be explicit, the ALASSO used the penalty form $\lambda \sum_{j=1}^{d} |\beta_j|\tau_j$, where the positive weights τ_j's are data-driven so that small weights are chosen for big coefficients and big weights for small coefficients. The tuning parameter λ controls the complexity of the model. We show that in theory if the weights τ_j's are chosen properly, the ALASSO estimator will have the oracle properties. We also suggest the practical choices of τ_j's for real problems.

2 Variable Selection Using Penalized Partial Likelihood

Suppose a random sample of n individuals is chosen. Let T_i and C_i be the failure time and censoring time of subject i $(i = 1, \cdots, n)$, respectively. Define $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$. We use z_{ij} to denote the jth covariate value for the ith individual, and $\mathbf{z}_i = (z_{i1}, \cdots, z_{id})^T$ to denote the vector of covariates for the ith individual. Assume that T_i and C_i are conditionally independent given $\mathbf{z}_i$, and the censoring mechanism is noninformative. The data then consists of the triplets $(\tilde{T}_i, \delta_i, \mathbf{z}_i),\ i = 1, ...n$.

The proportional hazards model (1) is assumed for the failure times T_i's. For simplicity, assume that there are no ties in the observed failure times. When ties are present, we may use the technique in Breslow (1974). The log partial likelihood is then given by

$$l_n(\boldsymbol{\beta}) \equiv \sum_{i=1}^{n} \delta_i \big\{ \boldsymbol{\beta}' \mathbf{z}_i - \log[\sum_{j=1}^{n} I(\tilde{T}_j \ge \tilde{T}_i) \exp(\boldsymbol{\beta}' \mathbf{z}_j)] \big\}. \tag{2}$$

To select important variables under the proportional hazards model, Tibshirani (1997) and

Fan & Li (2002) proposed to minimize the following penalized log partial likelihood function,

$$-\frac{1}{n}l_n(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{d} J(\beta_j). \tag{3}$$

Tibshirani (1997) suggested the L_1 penalty $J(\beta_j) = |\beta_j|$, leading to the so-called LASSO estimates. It is known that the L_1 penalty can shrink small coefficients to exactly zeros and hence result in a sparse representation of the solution. However, the LASSO applies the same penalty to all the coefficients. Large values of β's may suffer from substantial bias if λ is chosen too big, while the model may not be sufficiently sparse if λ is chosen too small. Fan & Li (2002) proposed the SCAD (smoothly clipped absolute deviation) penalty on $\boldsymbol{\beta}$, and the resulting estimator has nice theoretical properties. However, since the SCAD penalty is not convex in $\boldsymbol{\beta}$, its implementation can be challenging in practice. In the next section, we show how the adaptive LASSO improves the LASSO by using the data-driven penalties and hence achieves the theoretical properties of the SCAD.

3 Adaptive LASSO Estimation

We propose the adaptive Lasso (ALASSO) procedure, which solves the following penalized partial likelihood problem

$$\min_{\boldsymbol{\beta}} -\frac{1}{n}l_n(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{d} |\beta_j|\tau_j, \tag{4}$$

where the positive weights $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_d)'$ are chosen adaptively by data. The τ_j's can be regarded as leverage factors to adjust penalties on individual regression coefficients, taking large values for unimportant covariates and small values for important covariates. As we show later in Section 4, the choice of τ_j's is essential and their appropriate values will guarantee the optimality of the ALASSO solution. In the paper, we propose using $\tau_j =$

$1/|\tilde{\beta}_j|$, where $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \cdots, \tilde{\beta}_d)'$ is the maximizer of the log partial likelihood $l_n(\boldsymbol{\beta})$. Since $\tilde{\boldsymbol{\beta}}$ are consistent estimates (Tsiatis, 1981; Andersen & Gill, 1982) , their values well reflect the relative importance of covariates. In the paper we focus on the following minimization problem

$$\min_{\boldsymbol{\beta}} -\frac{1}{n} l_n(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{d} |\beta_j| / |\tilde{\beta}_j|. \tag{5}$$

Other consistent estimates of β's can be used as well; here $\tilde{\boldsymbol{\beta}}$ is just a convenient choice. We observe that the adaptive penalty term in (5) is closely related to the L_0 penalty $\sum_{j=1}^{d} I(|\beta_j| \neq 0)$, also called the entropy penalty in wavelet literature (Donoho & Johnstone 1998; Antoniadis & Fan 2001). Due to the consistency of $\tilde{\beta}_j$, the term $|\beta_j|/|\tilde{\beta}_j|$ converges to $I(\beta_j \neq 0)$ in probability as the sample size goes to infinity. Therefore the ALASSO procedure can be regarded as an automatic implementation of the best subset selection in some asymptotic sense.

3.1 Computational Algorithm

To solve (5), we approximate the partial likelihood function using the Newton-Raphson update through an iterative least square procedure similar to that of Tibshirani (1997), and at each iteration we solve the least squares subject to the weighted L_1 penalty. Define the gradient vector $\nabla l(\boldsymbol{\beta}) = -\partial l_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ and the Hessian matrix $\nabla^2 l(\boldsymbol{\beta}) = -\partial^2 l_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \boldsymbol{\beta}'$. Let X denote the Cholesky decomposition of $\nabla^2 l(\boldsymbol{\beta})$, i.e. $\nabla^2 l(\boldsymbol{\beta}) = X'X$, and set the pseudo response vector $\mathbf{Y} = (X')^{-1}(\nabla^2 l(\boldsymbol{\beta})\boldsymbol{\beta} - \nabla l(\boldsymbol{\beta}))$. By the second-order Taylor expansion, $-l_n(\beta)$ can be approximated by the quadratic form $\frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})'(\mathbf{Y} - X\boldsymbol{\beta})$. Thus at each iterative step, we minimize

$$\frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta})'(\mathbf{Y} - X\boldsymbol{\beta}) + \lambda \sum_{j=1}^{d} |\beta_j|/|\tilde{\beta}_j|. \tag{6}$$

For solving the standard LASSO (i.e. all the weights are equal to 1), Tibshirani (1996) suggested two algorithms based on quadratic programming techniques, and Fu (1998) proposed the shooting algorithm. Recently Efron et al. (2004) showed that, under the least squares setting, the whole solution path of LASSO can be obtained by a modified LARS algorithm. In this paper, we have modified Fu's shooting algorithm to take into account different weights in the ALASSO. For any fixed λ, we propose the following algorithm:

1. Solve $\tilde{\boldsymbol{\beta}}$ by minimizing the negative log partial likelihood $-l_n(\boldsymbol{\beta})$.

2. Initialization: $k = 1$ and $\hat{\boldsymbol{\beta}}_{[1]} = \mathbf{0}$.

3. Compute $\nabla l, \nabla^2 l, X, \mathbf{Y}$ based on the current value $\hat{\boldsymbol{\beta}}_{[k]}$.

4. Minimize (6) using the modified shooting algorithm. Denote the solution as $\hat{\boldsymbol{\beta}}_{[k+1]}$.

5. Let $k = k + 1$. Go back to step 3 until the convergence criterion meets.

3.2 Variance Estimation and Parameter Tuning

Assume the true parameter $\boldsymbol{\beta}_0 = \{(\boldsymbol{\beta}_0^{(1)})', (\boldsymbol{\beta}_0^{(2)})'\}'$, where $\boldsymbol{\beta}_0^{(1)}$ consists of nonzero components and $\boldsymbol{\beta}_0^{(2)}$ consists of zero components. Define $A(\boldsymbol{\beta}) = \mathrm{diag}\{1/\beta_1^2, \cdots, 1/\beta_d^2\}$,

$$D(\boldsymbol{\beta}) = \mathrm{diag}\left\{\frac{I(\beta_1 \neq 0)}{\beta_1^2}, \cdots, \frac{I(\beta_d \neq 0)}{\beta_d^2}\right\}, \text{ and } b(\boldsymbol{\beta}) = \left(\frac{\mathrm{sign}(|\beta_1|)}{|\tilde{\beta}_1|}, \cdots, \frac{\mathrm{sign}(|\beta_d|)}{|\tilde{\beta}_d|}\right)'.$$

At the $(k + 1)$th step, the ALASSO solution can be approximated by

$$\hat{\boldsymbol{\beta}}_{[k+1]} = \hat{\boldsymbol{\beta}}_{[k]} - \left[\nabla^2 l(\hat{\boldsymbol{\beta}}_{[k]}) + \lambda A(\hat{\boldsymbol{\beta}}_{[k]})\right]^{-1} \left[\nabla l(\hat{\boldsymbol{\beta}}_{[k]}) + \lambda b(\hat{\boldsymbol{\beta}}_{[k]})\right].$$

Using similar techniques in Fan & Li (2002), the covariance matrix of the ALASSO estimator $\hat{\boldsymbol{\beta}}$ can be approximated by the following sandwich formula:

$$\left\{\nabla^2 l(\hat{\boldsymbol{\beta}}) + \lambda A(\hat{\boldsymbol{\beta}})\right\}^{-1} \Sigma(\hat{\boldsymbol{\beta}}) \left\{\nabla^2 l(\hat{\boldsymbol{\beta}}) + \lambda A(\hat{\boldsymbol{\beta}})\right\}^{-1},$$

where $\Sigma(\boldsymbol{\beta}) = \{\nabla^2 l(\boldsymbol{\beta}) + \lambda D(\boldsymbol{\beta})\}\{\nabla^2 l(\boldsymbol{\beta})\}^{-1}\{\nabla^2 l(\boldsymbol{\beta}) + \lambda D(\boldsymbol{\beta})\}$. Write $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$, where $\hat{\boldsymbol{\beta}}_1$ consists of the r non-zero components. Correspondingly, we decompose the Hessian matrix as

$$G = \nabla^2 l(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where G_{11} corresponds to the first $r \times r$ submatrix of G. Similarly, let A_{11} be the first $r \times r$ submatrix of $A \equiv A(\hat{\boldsymbol{\beta}})$. Define $E = G_{22} - G_{21}G_{11}^{-1}G_{12}$ and $\tilde{G}_{11} = G_{11} + \lambda A_{11}$. It is easy to show that the covariance estimate of the nonzero components $\hat{\boldsymbol{\beta}}_1$ is

$$\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}_1) = G_{11}^{-1} + \left[G_{11}^{-1} - \tilde{G}_{11}^{-1}\right] G_{12}E^{-1}G_{21} \left[G_{11}^{-1} - \tilde{G}_{11}^{-1}\right].$$

To estimate the tuning parameter λ, we use the generalized cross validation (GCV) criterion (Craven & Wahba, 1979). At convergence, the minimizer of (6) in step 4 can be approximated by a ridge regression estimator $\hat{\boldsymbol{\beta}}_r = (G + \lambda A)^{-1}X'\mathbf{Y}$. Therefore the number of effective parameters in the ALASSO estimator can be approximated by $p(\lambda) = \text{tr}[(G + \lambda A)^{-1}G]$. The GCV-type statistic is constructed as $\text{GCV}(\lambda) = -l_n(\hat{\boldsymbol{\beta}})/[n\{1 - p(\lambda)/n\}^2]$.

4 Theoretical Properties of ALASSO Estimator

In this section, we study the asymptotic properties of the ALASSO estimator from two perspectives. Consider the penalized log partial likelihood function based on n samples

$$Q_n(\boldsymbol{\beta}) = l_n(\boldsymbol{\beta}) - n\lambda_n \sum_{j=1}^{d} |\beta_j|/|\tilde{\beta}_j|. \tag{7}$$

Denote the ALASSO solution of (7) by $\hat{\boldsymbol{\beta}}_n$. Recall $\boldsymbol{\beta}_0 = (\beta_{10}, ..., \beta_{d0}) = \{(\boldsymbol{\beta}_0^{(1)})', (\boldsymbol{\beta}_0^{(2)})'\}'$. Write $\hat{\boldsymbol{\beta}}_n = (\hat{\beta}_{1n}, ..., \hat{\beta}_{dn}) = \{(\hat{\boldsymbol{\beta}}_n^{(1)})', (\hat{\boldsymbol{\beta}}_n^{(2)})'\}'$ accordingly. Assume the length of the true nonzero components $\boldsymbol{\beta}_0^{(1)}$ of $\boldsymbol{\beta}_0$ is s. Let $I(\boldsymbol{\beta}_0)$ be the Fisher information matrix based on the log partial likelihood and $I_1(\boldsymbol{\beta}_0^{(1)}) = I_{11}(\boldsymbol{\beta}_0^{(1)}, \mathbf{0})$, where $I_{11}(\boldsymbol{\beta}_0^{(1)}, \mathbf{0})$ is the first $s \times s$ submatrix of $I(\boldsymbol{\beta}_0)$ knowing $\boldsymbol{\beta}_0^{(2)} = \mathbf{0}$. We first show that $\hat{\boldsymbol{\beta}}_n$ is root-n consistent if $\lambda_n \to 0$ with certain rate. Secondly, we show that the ALASSO estimator must satisfy $\hat{\boldsymbol{\beta}}_n^{(2)} = \mathbf{0}$ and $\hat{\boldsymbol{\beta}}_n^{(1)}$ is asymptotic normal with covariance matrix $I_1^{-1}(\boldsymbol{\beta}_0^{(1)})$ if $n\lambda_n \to \infty$. This is the so-called oracle property (Donoho & Johnstone, 1994), implying that the estimator works as well as if the correct submodel were known.

Define the counting and at-risk processes $N_i(t) = \delta_i I(\tilde{T}_i \leq t)$ and $Y_i(t) = I(\tilde{T}_i \geq t)$, respectively. In this section, the covariate $\mathbf{z}$ is allowed to be time-dependent, denoted by $\mathbf{z}(t)$. Without loss of generality, we assume $t \in [0, 1]$. Then the fisher information matrix is given by

$$I(\boldsymbol{\beta}_0) = \int_0^1 v(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, t) h_0(t) dt,$$

where

$$v(\boldsymbol{\beta}, t) = \frac{s^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \left(\frac{s^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}\right)\left(\frac{s^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}\right)',$$

and $s^{(k)}(\boldsymbol{\beta}, t) = E[\mathbf{z}(t)^{\otimes k} Y(t) \exp\{\boldsymbol{\beta}'\mathbf{z}(t)\}]$, $k = 0, 1, 2$. The regularity conditions (A) - (D) used in Anderson and Gill (1982) are assumed in the whole section.

Theorem 1. *(Consistency) Assume that* $(\mathbf{z}_1, T_1, C_1), ..., (\mathbf{z}_n, T_n, C_n)$ *are independently and identically distributed, and* T_i *and* C_i *are independent given* $\mathbf{z}_i$. *If* $\sqrt{n}\lambda_n = O_p(1)$, *then the ALASSO estimator satisfies* $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$.

Theorem 2. *(Oracle Property) Assume that* $\sqrt{n}\lambda_n \to \lambda_0$ *with* $0 \leq \lambda_0 < \infty$ *and* $n\lambda_n \to \infty$, *then under the conditions of Theorem 1, with probability tending to 1, the root-n consistent ALASSO estimator* $\hat{\boldsymbol{\beta}}_n$ *must satisfy:*

(i) (Sparsity) $\hat{\boldsymbol{\beta}}_n^{(2)} = \mathbf{0}$*;*

(ii) (Asymptotic normality) $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_0^{(1)}) \rightarrow N\{-\lambda_0 I_1^{-1}(\boldsymbol{\beta}_0^{(1)})\mathbf{b}_1, I_1^{-1}(\boldsymbol{\beta}_0^{(1)})\}$ *as n goes to infinity, where* $\mathbf{b}_1 = (sign(\beta_{10})/|\beta_{10}|, \cdots, sign(\beta_{s0})/|\beta_{s0}|)'$ *and* $\boldsymbol{\beta}_0^{(1)} = (\beta_{10}, \cdots, \beta_{s0})'$.

The proofs of Theorems 1 and 2 are given in the Appendix. When $\lambda_0 = 0$, the ALASSO estimator performs as well as the maximum partial likelihood estimates for estimating $\boldsymbol{\beta}_0^{(1)}$ knowing $\boldsymbol{\beta}_0^{(2)} = \mathbf{0}$. Note that the SCAD estimator (Fan & Li, 2002) also has above two properties under different regularity conditions for λ_n, but the LASSO estimator does not. Since the proofs only require the root-n consistency of $\tilde{\boldsymbol{\beta}}$, we want to emphasize that, any root-n consistent estimates of $\boldsymbol{\beta}_0$ can be used for the adaptive weights $\boldsymbol{\tau}$ without changing the asymptotic properties of the ALASSO solution. In addition to this theoretical improvement of the LASSO via the adaptive weighting scheme, in the next session we show that the ALASSO also demonstrates much better performance on numerical examples.

5 Numerical Studies

5.1 Simulations

Several simulation studies were conducted to assess the performance of the ALASSO, LASSO, and maximum partial likelihood estimates (MLE) under Cox's proportional hazards model. We report the average numbers of correct and incorrect zero coefficients in the final models. Following Tibshirani (1997), we measure the prediction error of each method with the mean of the estimated mean squared errors (MSE) $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T V(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ over 100 simulations. Here V is the population covariance matrix of the covariates. Generalized cross validation is used to estimate the tuning parameter λ in the ALASSO and LASSO methods. All simulations are done with R codes.

The failure times are generated from the proportional hazards model (1) with $\boldsymbol{\beta} = (-0.7, -0.7, 0, 0, 0, -0.7, 0, 0, 0)'$. The nine covariates $\mathbf{z} = (z_1, ..., z_9)$ are marginally standard normal with the pairwise correlation $\text{corr}(z_j, z_k) = \rho^{|j-k|}$. We consider the moderate correlation between the covariates with $\rho = 0.5$. Censoring times are generated from the uniform distribution over $[0, c_0]$, where c_0 is chosen to obtain the desired censoring rate. We consider two types of censoring rate: 25% and 40%, and two sample sizes: $n = 100$ and $n = 200$. Table 1 summarizes the mean square errors and variable selection results for three methods under four different settings. Overall, the ALASSO outperforms the other two approaches in terms of MSE and the correct number of zero coefficients in the solution. For example, when $n = 200$ and the censoring rate is 25%, the Alasso selects important covariates very accurately (the true model size is 3, MLE 8, LASSO 4.06, ALASSO 3.09), and gives the smallest mean squared error (MLE 0.097, LASSO 0.101, ALASSO 0.069). In Table 2, we show the number of times that each variable is selected among 100 replicates. The ALASSO chooses unimportant variables with a much lower frequency than the LASSO in all the settings.

(Insert Tables 1 and 2 here)

To test the accuracy of the proposed standard error formula given in Section 3·2, we compare the sample standard errors with their estimates. For the LASSO estimates, we use the formula in Tibshirani (1997) to compute their standard errors. In Table 3, we summarize the mean of the estimated standard errors and the sample standard errors from

Monte Carlo simulations. The estimated standard errors of all the three methods are close to the sample standard errors. And the performance becomes better when the sample size increases.

(Insert Table 3 here)

5.2 Primary Biliary cirrhosis Data

The primary biliary cirrhosis (PBC) data was gathered from the Mayo Clinic trial in primary biliary cirrhosis of liver conducted between 1974 and 1984. This data is provided in Therneau and Grambsch (2000), and a more detailed account can be found in Dickson *et al.* (1989). In this study, 312 patients from a total of 424 patients who agreed to participate in the randomized trial are eligible for the analysis. For each patient, clinical, biochemical, serologic, and histological parameters are collected. 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 for control, 2 for treatment).

We restrict our attention to the 276 observations without missing values. All the seventeen variables are included in the model. Table 4 summarizes the estimated coefficients by three methods and the corresponding standard errors. As reported in Tibshirani (1997), the stepwise selection chooses eight variables: *age, ede,bili, alb, cop, sgot, prot* and *stage.* We found that the ALASSO identifies the same set of important variables. In our analysis, the LASSO selects one additional variable *asc.*

(Insert Table 4 here)

6 Discussion

In this paper, we propose using the penalized partial likelihood with an adaptive L_1 penalty for model selection and estimation under Cox's hazard hazards models. The proposed ALASSO estimator has the oracle properties and is easy to solve with its convex penalty form. Numerical examples show that the ALASSO gives better prediction performance than the classical maximum partial likelihood estimate and selects more correct models than its Lasso variant.

For the ALASSO procedure, the choice of the weights τ_j's is very important. In the paper, we suggest using the inverse of the absolute maximum partial likelihood estimate, i.e. $\tau_j = 1/|\tilde{\beta}_j|$ for its convenience and nice properties. However, in some complicated situations, $\tilde{\beta}_j$'s may not be estimable, for example in high dimensional gene expression data where the number of covariates d is much large than the sample size n; or $\tilde{\beta}_j$'s may be unstably estimated if strong collinearity exists among covariates. In such cases, we suggest using some robust estimates such as ridge regression estimates for the weights.

Acknowledgment

Hao Helen Zhang's research was partially supported by National Science Foundation

Grant DMS-0405913. Wenbin Lu's research was partially supported by National Science Foundation Grant DMS-0504269.

Appendix

We follow the similar steps in the proofs of Fan & Li (2002).

A.1 Proof of Theorem 1

The log partial likelihood $l_n(\boldsymbol{\beta})$ can be written as

$$l_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^1 \boldsymbol{\beta}^T \mathbf{z}_i(s) dN_i(s) - \int_0^1 \log \left[\sum_{i=1}^{n} Y_i(s) \exp\{\boldsymbol{\beta}^T \mathbf{z}_i(s)\} \right] d\bar{N}(s), \tag{A1}$$

where $\bar{N} = \sum_{i=1}^n N_i$. By Theorem 4.1 and Lemma 3.1 of Anderson and Gill (1982), it follows that for each $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$,

$$\begin{aligned} \frac{1}{n}\{l_n(\boldsymbol{\beta}) - l_n(\boldsymbol{\beta}_0)\} &= \int_0^1 \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T s^{(1)}(\boldsymbol{\beta}_0, t) - \log \left\{ \frac{s^{(0)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\} s^{(0)}(\boldsymbol{\beta}_0, t) \right] \lambda_0(t) dt \\ &+ O_p(\frac{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|}{\sqrt{n}}). \end{aligned} \tag{A2}$$

Consider the C-ball $B_n(C) = \{\boldsymbol{\beta} : \boldsymbol{\beta} = \boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u}, ||\mathbf{u}|| \le C\}$, $C > 0$, and denote its boundary by $\partial B_n(C)$. Note that $Q_n(\boldsymbol{\beta})$ is strictly convex when n is large. Thus, there exists a unique maximizer $\hat{\boldsymbol{\beta}}_n$ of $Q_n(\boldsymbol{\beta})$ for large n. It is sufficient to show: for any given $\epsilon > 0$, there exists a large constant C so that

$$P\left\{ \sup_{\boldsymbol{\beta} \in \partial B_n(C)} Q_n(\boldsymbol{\beta}) < Q_n(\boldsymbol{\beta}_0) \right\} \ge 1 - \epsilon. \tag{A3}$$

This implies with probability at least $1-\epsilon$ that there exists a local maximizer of $Q_n(\boldsymbol{\beta})$ in the ball $B_n(C)$. Hence, the maximizer $\hat{\boldsymbol{\beta}}_n$ must satisfy $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$.

Define $s_n(\boldsymbol{\beta}) = \partial l_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ and $\bigtriangledown s_n(\boldsymbol{\beta}) = \partial s_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}'$. We have $s_n(\boldsymbol{\beta}_0)/\sqrt{n} = O_p(1)$ and $\bigtriangledown s_n(\boldsymbol{\beta}_0)/n = I(\boldsymbol{\beta}_0) + o_p(1)$. For any $\boldsymbol{\beta} \in \partial B_n(C)$, by the second-order Taylor expansion of the log partial likelihood

$$\begin{aligned}
&\frac{1}{n}\{l_n(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-l_n(\boldsymbol{\beta}_0)\}\\
=\;&\frac{1}{n}s_n'(\boldsymbol{\beta}_0)n^{-1/2}\mathbf{u}-\frac{1}{2n}\mathbf{u}'\{\bigtriangledown s_n(\boldsymbol{\beta}_0)/n\}\mathbf{u}+\frac{1}{n}\mathbf{u}'o_p(1)\mathbf{u}\\
=\;&-\frac{1}{2n}\mathbf{u}'\{I(\boldsymbol{\beta}_0)+o_p(1)\}\mathbf{u}+\frac{1}{n}O_p(1)\sum_{j=1}^{d}|u_j|,
\end{aligned}$$

where $\mathbf{u}=(u_1,\cdots,u_d)'$. Then we have

$$\begin{aligned}
D_n(\mathbf{u}) \;\equiv\;& \frac{1}{n}\{Q_n(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-Q_n(\boldsymbol{\beta}_0)\}\\
=\;&\frac{1}{n}\{l_n(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-l_n(\boldsymbol{\beta}_0)\}-\lambda_n\sum_{j=1}^{d}\left\{\frac{|\beta_{j0}+n^{-1/2}u_j|}{|\tilde{\beta}_j|}-\frac{|\beta_{j0}|}{|\tilde{\beta}_j|}\right\}\\
\le\;&\frac{1}{n}\{l_n(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-l_n(\boldsymbol{\beta}_0)\}-\lambda_n\sum_{j=1}^{s}(|\beta_{j0}+n^{-1/2}u_j|-|\beta_{j0}|)/|\tilde{\beta}_j|\\
\le\;&\frac{1}{n}\{l_n(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-l_n(\boldsymbol{\beta}_0)\}+n^{-1/2}\lambda_n\sum_{j=1}^{s}|u_j|/|\tilde{\beta}_j|\\
=\;&-\frac{1}{2n}\mathbf{u}'\{I(\boldsymbol{\beta}_0)+o_p(1)\}\mathbf{u}+\frac{1}{n}O_p(1)\sum_{j=1}^{d}|u_j|+\frac{1}{\sqrt{n}}\lambda_n\sum_{j=1}^{s}|u_j|/|\tilde{\beta}_j| \qquad \text{(A4)}
\end{aligned}$$

Using the fact that the maximum partial likelihood estimator $\tilde{\boldsymbol{\beta}}$ satisfies $||\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||=O_p(n^{-1/2})$, we have, for $1\le j\le s$,

$$\frac{1}{|\tilde{\beta}_j|}=\frac{1}{|\beta_{j0}|}-\frac{\mathrm{sign}(\beta_{j0})}{\beta_{j0}^2}(\tilde{\beta}_j-\beta_{j0})+o_p(|\tilde{\beta}_j-\beta_{j0}|)=\frac{1}{|\beta_{j0}|}+\frac{O_p(1)}{\sqrt{n}}.$$

In addition, since $\sqrt{n}\lambda_n=O_p(1)$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\lambda_n\sum_{j=1}^{s}|u_j|/|\tilde{\beta}_j| \;=\;& \frac{1}{\sqrt{n}}\lambda_n\sum_{j=1}^{s}\left\{\frac{|u_j|}{|\beta_{j0}|}+\frac{|u_j|}{\sqrt{n}}O_p(1)\right\}\\
\le\;& Cn^{-1/2}\lambda_nO_p(1)=Cn^{-1}(\sqrt{n}\lambda_n)O_p(1)=Cn^{-1}O_p(1).
\end{aligned}$$

Therefore in (A4), by choosing a sufficiently large C, the first term is of the order C^2n^{-1}. The second and third terms are of the order Cn^{-1}, which are dominated by the first term. Therefore (A3) holds and it completes the proof.

A.2 Proof of Theorem 2

(i) Show the sparsity: $\hat{\boldsymbol{\beta}}_n^{(2)} = \mathbf{0}$.

It is sufficient to show that for any sequence $\boldsymbol{\beta}_1$ satisfying that $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0^{(1)}\| = O_p(n^{-1/2})$ and any constant C,

$$Q_n(\boldsymbol{\beta}_1, \mathbf{0}) = \max_{\|\boldsymbol{\beta}_2\| \le Cn^{-1/2}} Q_n(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2).$$

It is sufficient to show that with probability tending to 1, for any $\boldsymbol{\beta}_1$ satisfying that $\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0^{(1)}\| = O_p(n^{-1/2})$, $\partial Q(\boldsymbol{\beta})/\partial \beta_j$ and β_j have different signs for $\beta_j \in (-Cn^{-1/2}, Cn^{-1/2})$ with $j = s+1, \cdots, d$. For each $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, by (A1) and Taylor expansion,

$$l_n(\boldsymbol{\beta}) = l_n(\boldsymbol{\beta}_0) + nf(\boldsymbol{\beta}) + O_P(\sqrt{n}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|),$$

where $f(\boldsymbol{\beta}) = -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\{I(\boldsymbol{\beta}_0) + o(1)\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. For $j = s+1, \cdots, d$, we have

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial l_n(\boldsymbol{\beta})}{\partial \beta_j} - n\lambda_n \frac{\mathrm{sign}(\beta_j)}{|\tilde{\beta}_j|} = O_p(n^{1/2}) - (n\lambda_n)n^{1/2}\frac{\mathrm{sign}(\beta_j)}{|n^{1/2}\tilde{\beta}_j|}.$$

Note that $n^{1/2}(\tilde{\beta}_j - 0) = O_p(1)$, we have

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \beta_j} = n^{1/2}\left\{O_p(1) - n\lambda_n \frac{\mathrm{sign}(\beta_j)}{|O_p(1)|}\right\}. \tag{A5}$$

Since $n\lambda_n \to \infty$, the sign of $\frac{\partial Q_n(\beta_j)}{\partial \beta_j}$ in (A5) is completely determined by the sign of β_j when n is large, and they always have different signs.

(ii) Show the asymptotic normality of $\hat{\boldsymbol{\beta}}_n^{(1)}$.

Using the proof of Theorem 1, it is easy to show that there exists a root-n consistent maximizer $\hat{\boldsymbol{\beta}}_n^{(1)}$ of $Q_n(\boldsymbol{\beta}_1, \mathbf{0})$, i.e. $\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}|_{\boldsymbol{\beta} = \{(\hat{\boldsymbol{\beta}}_n^{(1)})', \mathbf{0}'\}'} = \mathbf{0}$. Let $s_n^{(1)}(\boldsymbol{\beta})$ be the first s elements of $s_n(\boldsymbol{\beta})$ and $\hat{I}_n^{(11)}(\boldsymbol{\beta})$ be the first $s \times s$ submatrix of $\bigtriangledown s_n(\boldsymbol{\beta})$. Then

$$\begin{aligned} 0 &= \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}|_{\boldsymbol{\beta}=\{(\hat{\boldsymbol{\beta}}_n^{(1)})',\mathbf{0}'\}'} \\ &= \frac{\partial l_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1|}|_{\boldsymbol{\beta}=\{(\hat{\boldsymbol{\beta}}_n^{(1)})',\mathbf{0}'\}'} - n\lambda_n \left(\frac{\mathrm{sign}(\hat{\beta}_{1n})}{\tilde{\beta}_1}, \cdots, \frac{\mathrm{sign}(\hat{\beta}_{sn})}{\tilde{\beta}_s} \right)' \\ &= s_n^{(1)}(\boldsymbol{\beta}_0) - \hat{I}_n^{(11)}(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_0^{(1)}) - n\lambda_n \left(\frac{\mathrm{sign}(\beta_{10})}{\tilde{\beta}_1}, \cdots, \frac{\mathrm{sign}(\beta_{s0})}{\tilde{\beta}_s} \right)', \end{aligned}$$

where $\boldsymbol{\beta}^*$ is between $\hat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_0$. The last equation is implied by $\mathrm{sign}(\hat{\beta}_{jn}) = \mathrm{sign}(\beta_{j0})$ when n is large, since $\hat{\boldsymbol{\beta}}_n$ is a root-n consistent estimator of $\boldsymbol{\beta}_0$. Using Theorem 3.2 of Anderson and Gill (1982), it can be proved that $s_n^{(1)}(\boldsymbol{\beta}_0)/\sqrt{n} \to N\{\mathbf{0}, I_1(\boldsymbol{\beta}_0^{(1)})\}$ in distribution and $\hat{I}_n^{(11)}(\boldsymbol{\beta}^*)/n \to I_1(\boldsymbol{\beta}_0^{(1)})$ in probability as $n \to \infty$. In addition, $\sqrt{n}\lambda_n \to \lambda_0$ and $\tilde{\beta}_j \to \beta_{j0} \neq 0$, for $1 \leq j \leq s$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_0^{(1)}) = I_1^{-1}(\boldsymbol{\beta}_{10}) \left\{ \frac{1}{\sqrt{n}} s_n^{(1)}(\boldsymbol{\beta}_0) - \lambda_0 \mathbf{b}_1 \right\} + o_p(1).$$

Therefore, by Slutsky's Lemma,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_0^{(1)}) \to N\{-\lambda_0 I_1^{-1}(\boldsymbol{\beta}_0^{(1)})\mathbf{b}_1, I_1^{-1}(\boldsymbol{\beta}_0^{(1)})\}$$

in distribution as $n \to \infty$.

References

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study (Com: p1121-1124). *The Annals of Statistics* **10**, 1100–20.

Antoniadis, A. & Fan, J. (2001). Regularization of wavelets approximations. *Journal of the American Statistical Association* **96**, 939–63.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* **34**, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377–403.

Dickson, E., Grambsch, P., Fleming, T., Fisher, L., & Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* **10**, 1–7.

Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–55.

Donoho, D. L. & Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics* **26**, 879–921.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics* **32**, 407–51.

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

Faraggi, D. & Simon, R. (1998). Bayesian variable selection method for censored survival data. *Biometrics* **54**, 1475–85.

Fu, W. (1998). Penalized regression: the bridge versus the lasso. *Journal of Computational and Graphical Statistics* **7**, 397–416.

Ibrahim, J. G., Chen, M.-H., & MacEachern, S. N. (1999). Bayesian variable selection for proportional hazards models. *The Canadian Journal of Statistics* **27**, 701–17.

Sauerbrei, W. & Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the cox regression model. *Statistics in Medicine* **11**, 2093–109.

Therneau, T. M. & Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model*. Springer-Verlag Inc.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, Methodological* **58**, 267–88.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine* **16**, 385–95.

Tsiatis, A. (1981). A large sample study of cox's regression model. *The Annals of Statistics* **9**, 93–108.

Table 1: Mean squared error and model selection results.

		25% Censored			40% Censored		
n		Correct (6)	Incorrect (0)	MSE	Correct (6)	Incorrect (0)	MSE
	MLE	0	0	0.247	0	0	0.312
100	LASSO	4.87	0.00	0.190	4.67	0.00	0.203
	ALASSO	**5.73**	0.01	0.157	**5.63**	0.04	0.172
	MLE	0.00	0.00	0.097	0.00	0.00	0.113
200	LASSO	4.94	0.00	0.101	4.69	0.00	0.113
	ALASSO	**5.91**	0.00	0.069	**5.86**	0.00	0.074

Table 2: The frequency of variables selected by LASSO and ALASSO in 100 runs.

	n	Censored	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
	100	25%	100	100	29	15	16	100	16	15	12
LASSO		40%	100	100	27	22	22	100	23	11	18
	200	25%	100	100	14	16	24	100	19	17	16
		40%	100	100	21	23	31	100	21	18	17
	100	25%	99	100	5	7	3	100	2	4	6
ALASSO		40%	98	99	8	5	5	99	7	5	7
	200	25%	100	100	1	1	2	100	2	3	0
		40%	100	100	2	3	2	100	2	3	2

Table 3: Estimated and actual standard errors for the coefficients.

			$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_6$	
n	Censored		SE	$\widehat{\text{SE}}$	SE	$\widehat{\text{SE}}$	SE	$\widehat{\text{SE}}$
100	25%	MLE	0.192	0.169	0.214	0.186	0.216	0.185
		LASSO	0.154	0.105	0.153	0.104	0.158	0.096
		ALASSO	0.206	0.155	0.201	0.155	0.175	0.138
	40%	MLE	0.197	0.189	0.227	0.208	0.257	0.208
		LASSO	0.166	0.114	0.161	0.114	0.178	0.104
		ALASSO	0.218	0.170	0.211	0.171	0.208	0.151
200	25%	MLE	0.119	0.111	0.121	0.121	0.148	0.123
		LASSO	0.109	0.081	0.096	0.081	0.116	0.075
		ALASSO	0.128	0.107	0.116	0.106	0.131	0.096
	40%	MLE	0.133	0.123	0.136	0.134	0.152	0.136
		LASSO	0.124	0.089	0.113	0.088	0.131	0.082
		ALASSO	0.141	0.118	0.128	0.117	0.139	0.106

Table 4: Estimated coefficients and standard errors for PBC data.

Covariate	MLE	LASSO	ALASSO
trt	-0.124 (0.215)	0 (-)	0 (-)
age	0.029 (0.012)	0.015 (0.004)	0.019 (0.010)
sex	-0.366 (0.311)	0 (-)	0 (-)
asc	0.088 (0.387)	0.107 (0.052)	0 (-)
hep	0.026 (0.251)	0 (-)	0 (-)
spid	0.101 (0.244)	0 (-)	0 (-)
ede	1.011 (0.394)	0.648 (0.177)	0.671 (0.377)
bil	0.080 (0.025)	0.084 (0.013)	0.095 (0.020)
chol	0.001 (0.000)	0 (-)	0 (-)
alb	-0.742 (0.308)	-0.548 (0.133)	-0.612 (0.280)
cop	0.003 (0.001)	0.003 (0.001)	0.002 (0.001)
alk	0.000 (0.000)	0 (-)	0 (-)
sgot	0.004 (0.002)	0.001 (0.000)	0.001 (0.000)
trig	-0.001 (0.001)	0 (-)	0 (-)
plat	0.001 (0.001)	0 (-)	0 (-)
prot	0.233 (0.106)	0.125 (0.040)	0.103 (0.108)
stage	0.455 (0.175)	0.265 (0.064)	0.367 (0.142)