HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2017 December 22.

Published in final edited form as:

Biometrics. 2016 September ; 72(3): 731–741. doi:10.1111/biom.12464.

Model Selection and Inference for Censored Lifetime Medical Expenditures

Brent A. Johnson[1], **Qi Long**[2], **Yijian Huang**[2], **Kari Chansky**[3], and **Mary Redman**[3]

[1]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, U.S.A

[2]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, U.S.A

[3]The Fred Hutchinson Cancer Research Center, Seattle, Washington U.S.A

Summary

Identifying factors associated with increased medical cost is important for many micro- and macro-institutions, including the national economy and public health, insurers and the insured. However, assembling comprehensive national databases that include both the cost and individual-level predictors can prove challenging. Alternatively, one can use data from smaller studies with the understanding that conclusions drawn from such analyses may be limited to the participant population. At the same time, smaller clinical studies have limited follow-up and lifetime medical cost may not be fully observed for all study participants. In this context, we develop new model selection methods and inference procedures for secondary analyses of clinical trial data when lifetime medical cost is subject to induced dependent censoring. Our model selection methods extend a theory of penalized estimating function to a calibration regression estimator tailored for this data type. Next, we develop a novel inference procedure for the unpenalized regression estimator using perturbation and resampling theory. Then, we extend this resampling plan to accommodate regularized coefficient estimation of censored lifetime medical cost and develop post-selection inference procedures for the final model. Our methods are motivated by data from Southwest Oncology Group Protocol 9509, a clinical trial of patients with advanced nonsmall cell lung cancer, and our models of lifetime medical cost are specific to this population. But the methods presented in this article are built on rather general techniques and could be applied to larger databases as those data become available.

Keywords

Induced censoring; Marked point process; Regularization; Survival analysis

1. Introduction

Many institutions, including governments, hospitals, and private businesses, have great interest in factors associated with increased medical costs. For example, countries with state-

Supplementary Materials: Web Appendices A–C referenced in Sections 2, 3, and 4 are available with this paper at the Biometrics website on Wiley Online Library.

run medical insurance programs aim to provide state-of-the-art medicine to their citizens while moderating the financial burden on the economy. In our motivating study, the Southwest Oncology Group (SWOG) clinical trial 9509, Huang (2002) showed that while patients following either one of two competing treatment regimens for nonsmall cell lung cancer had similar survival, one treatment led to significantly lower medical costs compared to the other treatment. In this paper, we seek to identify systematically other predictors associated with differential medical costs in SWOG 9509, estimate their association and draw valid statistical inference on the final estimates. The analysis is complicated because clinical studies have limited follow-up and lifetime medical costs is not fully observed for all study participants.

Suppose Y is lifetime medical cost measured at the end of life, say at time T, and $\mathbf{z} = (z_1, \ldots, z_d)^{\mathrm{T}}$ is a d-vector of independent regressors. We assume the natural logarithm of lifetime medical cost is linearly related to the independent variables though the regression model,

$$\log Y = \sum_{j=1}^{d} z_j \beta_j + \varepsilon_Y, \quad (1)$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^{\mathrm{T}}$ are unknown coefficients and the error ε_Y comes from an unknown distribution. The primary scientific interest lies in the coefficient vector $\boldsymbol{\beta}$ but estimation is challenging because lifetime cost Y is observed only for uncensored observations, i.e. $T \leq C$, where C be a censoring random variable. It is well-known that naïve estimators of $\boldsymbol{\beta}$, e.g. least squares estimation on uncensored medical costs, work only under restrictive assumptions on the data-generating mechanism and are not, in general, consistent for $\boldsymbol{\beta}$ (e.g. Lin et al., 1997; Huang and Louis, 1998; Huang, 2002; Jain and Strawderman, 2002). Thus, the statistical objective is to estimate $\boldsymbol{\beta}$ consistently under reasonably general conditions with the observed data $(\delta Y, X, \delta, \mathbf{z})$, where $X = \min(T, C)$ and $\delta = I(T \leq C)$.

A principal challenge in estimating $\boldsymbol{\beta}$ in the presence of censoring is the potential non-identifiability of the conditional distribution $[Y|\mathbf{z}]$ when $\delta = 1$ (e.g. Zhao and Tsiatis, 1997; Huang and Louis, 1998). Two estimation strategies that accommodate both the identifiability of $[Y|\mathbf{z}]$ and the induced censoring include inverse weighting (Zhao and Tsiatis, 1997; Lin, 2000; Bang and Tsiatis, 2000; Jain and Strawderman, 2002; Tsiatis, 2006) and bivariate modeling of time and cost together (Huang, 2002). Due to space limitations, a detailed review of these strategies is not given here and we refer the interested reader elsewhere in the literature (Huang, 2009). Briefly, the inverse weighting strategy overcomes identifiability concerns by redefining the outcome as time-restricted cost such that the new outcome is identified by definition and overcomes censoring by weighting the uncensored observations in an uncensored least-squares estimator, for example, by the inverse probability of censoring via $P(C > t|\mathbf{z})$, $t \geq 0$. Alternatively, Huang (2002) suggested modeling cost and time together, for example, by parameterizing the conditional distribution $[(Y, T)|\mathbf{z}]$ as $[T|\mathbf{z}]$ and $[Y|(T, \mathbf{z})]$. Compared with inverse weighting, Huang's (2002) technique models cost outcome Y rather than time-restricted medical cost and estimates nuisance parameters through the conditional distribution $[T|\mathbf{z}]$ rather than $[C|\mathbf{z}]$. Nevertheless, both estimation

strategies regard cost-scale inference as the primary analytic goal and other aspects of the statistical model as secondary.

Our estimator is based on an extension and application of a theory of penalized estimating function applied to Huang's (2002) calibration regression estimator. If $\mathbb{S}_\beta(\boldsymbol{\beta})$ is a consistent estimating function for $\boldsymbol{\beta}$, then Johnson et al. (2008) defined the penalized estimating function

$$\mathbb{S}_{\beta,\lambda}(\boldsymbol{\beta})=\mathbb{S}_\beta(\boldsymbol{\beta})-n\mathbf{q}_\lambda(|\boldsymbol{\beta}|)\mathrm{sgn}(\boldsymbol{\beta}), \quad (2)$$

where $\mathrm{sgn}(\boldsymbol{\beta}) = (\mathrm{sign}(\beta_j), j = 1, \ldots, d)$, $\mathbf{q}_\lambda(|\boldsymbol{\beta}|)\mathrm{sgn}(\boldsymbol{\beta})$ is the element-wise product and $\mathbf{q}_\lambda(|\boldsymbol{\beta}|)$ is a penalty function satisfying conditions that we state in Section 2. A root to the estimating function $\mathbb{S}_{\beta,\lambda}(\boldsymbol{\beta})$ is both sparse and satisfies an oracle property (Fan and Li, 2001; Zou, 2006; Johnson et al., 2008). In our medical cost application, there are secondary parameters that must be estimated and, naturally, one might expect that different estimation strategies at this level will have consequences on the cost-scale estimator. We investigate the operating characteristics of two different approaches in this paper.

Variable selection procedures for censored medical cost are given in Section 2. This includes a method that estimates secondary parameters via unregularized estimation and another method that attempts to select variables and estimate parameters in $[Y|(T, \mathbf{z})]$ and $[T|\mathbf{z}]$ simultaneously; the former method is our preferred approach and was detailed in an unpublished technical report (Johnson et al., 2012) while the latter method is new and given for comparison purposes. New inference procedures for the unregularized and regularized calibration estimators are given in Section 3. For the regularized case, we adapt the post-selection inference procedures by Minnier et al. (2011) to calibration estimator for lifetime medical cost. We re-analyze the medical cost data from SWOG 9509 in Section 4 and report on our simulation studies in Section 5.

2. Variable Selection for Lifetime Medical Cost

2.1 Regularization in Cost-scale Only

Using a theory of counting processes (Andersen et al., 1993; Kalbfleisch and Prentice, 2002), the weighted log-rank estimating function for inference in the accelerated lifetime model (Tsiatis, 1990; Wei et al., 1990; Ying, 1993), and results for marked processes (Huang and Louis, 1998), Huang (2002) proposed the coefficient estimator for $(\boldsymbol{\beta}, \boldsymbol{\vartheta})$ in the bivariate regression model

$$\log Y=\sum_{j=1}^{d} z_j\beta_j+\varepsilon_Y, \quad \log T=\sum_{j=1}^{d} z_j\vartheta_j+\varepsilon_T, \quad (3)$$

as a root to the system of equations, $o_p(n^{1/2}) = \mathbb{S}(\boldsymbol{\theta}) \equiv \{\mathbb{S}_\beta(\boldsymbol{\theta}), \mathbb{S}_\vartheta(\boldsymbol{\theta})\}$,

$$\mathbb{S}_{\beta}(\boldsymbol{\theta})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\gamma(u,\boldsymbol{\vartheta})\{\mathbf{z}_i-\bar{\mathbf{z}}(u,\boldsymbol{\vartheta})\}\psi(\log Y_i-\mathbf{z}_i^T\boldsymbol{\beta})\mathrm{dN}_i(u,\boldsymbol{\vartheta}), \quad (4)$$

$$\mathbb{S}_{\vartheta}(\boldsymbol{\theta})=\sum_{i=1}^{n}\int_{-\infty}^{\infty}\gamma(u,\boldsymbol{\vartheta})\{\mathbf{z}_i-\bar{\mathbf{z}}(u,\boldsymbol{\vartheta})\}\mathrm{dN}_i(u,\boldsymbol{\vartheta}), \quad (5)$$

where $(\delta_i Y_i, X_i, \delta_i, \mathbf{z}_i)$ is the observed data for the i-th individual, $\gamma(t, \boldsymbol{\vartheta})$ is a non-negative weight function, $\psi(\cdot)$ is a strictly monotone function, $R_i(u,\vartheta)=I(\log X_i-\mathbf{z}_i^{\mathrm{T}}\vartheta\geq u)$ is the at-risk process, $N_i(u,\vartheta)=I(\log X_i-\mathbf{z}_i^{\mathrm{T}}\vartheta\leq u,\delta_i=1)$ is the counting process, and $\bar{\mathbf{z}}(u, \boldsymbol{\vartheta}) = \Sigma_j \mathbf{z}_j R_j(u, \boldsymbol{\vartheta})/\Sigma_j R_j(u, \boldsymbol{\vartheta})$. The statistic $\mathbb{S}_{\vartheta}(\boldsymbol{\theta})$ is the weighted log-rank estimating function (Tsiatis, 1990; Wei et al., 1990) and does not depend on the cost-scale coefficients $\boldsymbol{\beta}$. For arbitrary weight function $\gamma(u, \boldsymbol{\vartheta})$, the estimating function $\mathbb{S}_{\vartheta}(\boldsymbol{\theta})$ is known to be non-monotone, contain multiple roots, some of which may be inconsistent (Fygenson and Ritov, 1994). Under conditions (A)–(E) in Huang (2002), the joint coefficient estimator $\hat{\boldsymbol{\theta}}_0=(\hat{\boldsymbol{\beta}}_0^{\mathrm{T}},\hat{\boldsymbol{\vartheta}}_0^{\mathrm{T}})^{\mathrm{T}}$ is consistent and asymptotically normal with mean $\boldsymbol{\theta}_0$ and covariance $n^{-1}\{\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Gamma}^{-1})^{\mathrm{T}}\}$, where $\boldsymbol{\Gamma} = \nabla\{\lim_n n^{-1}\mathbb{S}(\boldsymbol{\theta}_0)\}$, $n^{-1/2}\mathbb{S}(\boldsymbol{\theta}_0) \rightarrow_d N(0, \boldsymbol{\Omega})$, and $(\boldsymbol{\theta})=\{{}_{\beta}^{\mathrm{T}}(\boldsymbol{\theta}),{}_{\vartheta}^{\mathrm{T}}(\boldsymbol{\theta})\}^{\mathrm{T}}$. The asymptotic covariance is not directly estimable because the asymptotic slope matrix $\boldsymbol{\Gamma}$ depends on the hazard functions of the errors $(\varepsilon_Y, \varepsilon_T)$ in (3). This point makes statistical inference for $\hat{\boldsymbol{\theta}}_0$, and hence $\hat{\boldsymbol{\beta}}_0$, challenging.

The calibration estimator $\hat{\boldsymbol{\beta}}_0$ consistently estimates the mark-scale regression coefficients $\boldsymbol{\beta}$ in (3) but does not perform variable selection. To simultaneously select and estimate cost-scale coefficients, we propose the system of estimating functions,

$$\mathbb{S}_{\lambda}(\boldsymbol{\theta})=\{{}_{\beta,\lambda}^{\mathrm{T}}(\boldsymbol{\theta}),{}_{\vartheta}^{\mathrm{T}}(\boldsymbol{\theta})\}, \quad (6)$$

where $\mathbb{S}_{\beta,\lambda}(\boldsymbol{\theta}) = \mathbb{S}_{\beta}(\boldsymbol{\theta}) - n\mathbf{q}_{\lambda}(|\boldsymbol{\beta}|)\mathrm{sgn}(\boldsymbol{\beta})$ is the penalized estimating function, analogous to (2), and $\mathbf{q}_{\lambda}(|\boldsymbol{\beta}|) = (q_{\lambda,1}(|\beta_1|), \ldots, q_{\lambda,d}(|\beta_d|))^{\mathrm{T}}$ satisfies the following two conditions: for fixed $\beta > 0$,

A1. $\lim_{n\rightarrow\infty} n^{1/2}q_{\lambda}(|\beta|) = 0$ and $\lim_{n\rightarrow\infty}(\partial/\partial\beta)q_{\lambda}(|\beta|) = 0$;

A2. For any $K > 0$, $\lim_{n\rightarrow\infty} n^{1/2} \inf_{|\beta|<Kn^{-1/2}} q_{\lambda}(|\beta|) \rightarrow \infty$.

Conditions A1–A2 are sufficient to define a cost-scale coefficient estimator, say $\hat{\boldsymbol{\beta}}_{\lambda}$, that achieves an oracle property. In words, the oracle property implies the coefficient estimator is consistent and asymptotically normal, and sets the coefficient estimate exactly to zero for unimportant variables with probability tending to one. The asymptotic properties of $\hat{\boldsymbol{\beta}}_{\lambda}$ are given in Web Appendix A. Note, the system of equations $\mathbb{S}_{\lambda}(\boldsymbol{\theta})$ regularizes the cost-scale

estimating function but not the time-scale estimating function; we revisit this point in Section 2.2.

The role of $\psi(\cdot)$ in (4) is to moderate the influence of extremely large observations on the mark-scale. But setting $\psi(y) = y$ leads to a computationally efficient estimator and an exact solution to the system of penalized estimating functions. In the sequel, the identity weight function $\psi(\cdot)$ is assumed throughout. Then, one can show that the penalized estimating function $\mathbb{S}_{\beta,\lambda}(\boldsymbol{\beta}, \boldsymbol{\vartheta}_0)$ in $\mathbb{S}_\lambda(\boldsymbol{\theta})$ is the quasi-gradient of the objective function,

$$\mathbb{Q}_{\beta,\lambda}(\boldsymbol{\beta}, \boldsymbol{\vartheta}_0) = \frac{1}{2}\{\mathbf{V}(\boldsymbol{\vartheta}_0) - \mathbf{A}(\boldsymbol{\vartheta}_0)\boldsymbol{\beta}\}^{\mathrm{T}}\{\mathbf{V}(\boldsymbol{\vartheta}_0) - \mathbf{A}(\boldsymbol{\vartheta}_0)\boldsymbol{\beta}\} + n\sum_{j=1}^{d} p_{\lambda,j}(\beta_j), \quad (7)$$

with $(\partial/\partial\beta_j)p_{\lambda,j}(\beta_j) = q_{\lambda,j}(|\beta_j|)\mathrm{sign}(\beta_j)$, $\mathbf{V}(\boldsymbol{\vartheta}) = \{\mathbf{A}^{\mathrm{T}}(\boldsymbol{\vartheta})\}^{-1}\mathbf{w}(\boldsymbol{\vartheta})$, $\mathbf{A}(\boldsymbol{\vartheta})$ is the Choleski decomposition of $\mathbf{M}(\boldsymbol{\vartheta})$, and $\mathrm{w}(\vartheta) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\gamma(u,\vartheta)Y_i\{\mathbf{z}_i - \bar{\mathbf{z}}(u,\vartheta)\}\mathrm{dN}_i(u,\vartheta)$, and $\mathrm{M}(\vartheta) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\gamma(u,\vartheta)\{\mathbf{z}_i - \bar{\mathbf{z}}(u,\vartheta)\}\mathbf{z}_i^{\mathrm{T}}\mathrm{dN}_i(u,\vartheta)$. The expression $\mathbb{Q}_{\beta,\lambda}(\boldsymbol{\beta}, \boldsymbol{\vartheta}_0)$ is a surrogate loss function depending on the unknown time-scale coefficient $\boldsymbol{\vartheta}_0$. Thus, $\hat{\boldsymbol{\beta}}_\lambda$ minimizes $\mathbb{Q}_{\beta,\lambda}(\boldsymbol{\beta}, \hat{\boldsymbol{\vartheta}}_0)$.

2.2 Simultaneous Regularization on Cost- and Time-scale Coefficients

An alternative approach to model selection for censored medical cost data would be to perform estimation and variable selection over cost- and time-scale coefficients simultaneously. First, there may be a scientific reason to regard both time- and cost-scale coefficients as primary parts of the statistical model. Second, one might hope to achieve a better cost-scale coefficient estimator by removing unimportant time-scale regressors or otherwise regularizing the time-scale parameters. We show one method of joint regularization using effectively the same techniques as in Section 2.1. By replacing $\boldsymbol{\beta}$ with $\boldsymbol{\theta}$ in (6), we define the penalized estimating function

$$\tilde{\mathbb{S}}_\lambda(\boldsymbol{\theta}) = (\boldsymbol{\theta}) - n\mathbf{q}_\lambda(|\boldsymbol{\theta}|)\mathrm{sgn}(\boldsymbol{\theta}), \quad (8)$$

where $\mathbf{q}_\lambda(\cdot)$ is defined through Conditions A1–A2. Under regularity conditions in Huang (2002, Appendix), Theorem 1 in Johnson et al. (2008) says that a root to $\tilde{\mathbb{S}}_\lambda(\boldsymbol{\theta})$ possesses an oracle property. Hence, both approaches via (6) and (8) result in sparse coefficient estimates on the cost-scale but only (8) results in sparse estimates on the time-scale.

Unfortunately, solving $\tilde{\mathbb{S}}_\lambda(\boldsymbol{\theta})$ using ordinary methods, e.g. local quadratic approximation, is difficult because $\boldsymbol{\Gamma}$ is not directly estimable. Wang and Leng (2007) proposed an indirect, but asymptotically equivalent, method to finding roots of penalized estimating functions. It is easy to show that under standard conditions on parametric families of distributions, a local approximation to the likelihood is the quadratic function $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}}\mathcal{I}_n(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, where $\mathcal{I}_n$ is the information matrix evaluated at $\boldsymbol{\theta}_0$. Under technical regularity assumptions on estimating

functions $\mathbb{S}(\boldsymbol{\theta})$, including the asymptotic linearity of $\mathbb{S}(\boldsymbol{\theta})$, Wang and Leng (2007) showed this idea extends to general *M*- and *Z*-estimators by replacing $\mathcal{Q}_n$ with the inverse sandwich matrix. Thus, a penalized local quadratic loss function for our problem is

$$\tilde{\mathbb{Q}}_\lambda(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \sum^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \sum_{j=1}^{d} p_{\lambda,j}(\theta_j), \tag{9}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Gamma}^{-1})^{\mathrm{T}}$ and $p_{\lambda,j}$ were defined in (7). In practice, $\boldsymbol{\Sigma}$ is unknown and must be estimated; see Section 3.

2.3 Algorithmic and Practical Notes

In the sequel, we consider two weight functions, the Gehan weight function, $\gamma(u, \vartheta) = n^{-1}\sum_{i=1}^{n} R_i(u, \vartheta)$ and the log-rank weight function, $\gamma(u, \vartheta) = 1$. To compute the Gehan estimate of the time-scale coefficient ϑ, we use the linear programming technique by Jin et al. (2003) to compute the Gehan estimates. Then, we compute the log-rank estimate through their iteratively reweighted Gehan estimate using 5 iterations. The unregularized mark-scale estimate is the solution to a linear system (Huang, 2002). To compute the regularized mark-scale estimates for general penalty function, Johnson et al. (2012) investigated several algorithms for non-concave penalized least squares. Due to space limitations, we refer interested readers to our technical report for details. The results in Sections 4-5 use a multi-stage local linear approximation algorithm.

The regularization parameter λ is determined by minimizing data-dependent information criteria. We define the quadratic function $D(\boldsymbol{\theta}) = n^{-1\mathrm{T}}(\boldsymbol{\theta})\boldsymbol{\Omega}_n^{-1}(\hat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta})$, where $\boldsymbol{\Omega}_n(\boldsymbol{\theta}_0)$ is a consistent estimator of the asymptotic covariance matrix $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$,

$$\boldsymbol{\Omega}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{ \begin{array}{c} \gamma(u,\vartheta)\psi(\log Y_i - \mathbf{z}_i^{\mathrm{T}}\beta) \\ \gamma(u,\vartheta) \end{array} \right\}^{\otimes 2} \otimes \{\mathbf{z}_i - \overline{\mathbf{z}}(u,\vartheta)\}^{\otimes 2} \mathrm{dN}_i(u,\vartheta),$$

$\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^{\mathrm{T}}$ for the vector $\mathbf{v}$ and $\mathbf{A} \otimes \mathbf{B}$ is the kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$. Following similar arguments to Wei et al. (1990), one can show that $D(\hat{\boldsymbol{\theta}}_\lambda)$ converges to a χ^2 random variable as $n \to \infty$ with degrees of freedom equal to the number of zero coefficients in $\hat{\boldsymbol{\theta}}_\lambda$. Thus, our BIC-type criterion is $\mathrm{BIC}(\lambda) = D(\hat{\boldsymbol{\theta}}_\lambda) + \log(n)\hat{d}(\lambda)$, where $\hat{d}(\lambda)$ is the cardinality of $\hat{\boldsymbol{\theta}}_\lambda$, and an AIC-type criterion is $\mathrm{AIC}(\lambda) = D(\boldsymbol{\theta}_\lambda) + 2\hat{d}(\lambda)$.

3. Inference Procedures

Similar to Minnier et al. (2011), our resampling procedure is based on perturbation theory. Let $(\zeta_1, \ldots, \zeta_n)$ be independent and identically distributed random variables, completely independent of the observed data, such that $E(\zeta_1) = \mathrm{var}(\zeta_1) = 1$. Without loss of generality, let $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\vartheta}}^*$ be resampled estimates for $\boldsymbol{\beta}$ and ϑ, respectively. Suppose that $\hat{\boldsymbol{\beta}}$ has influence curve $\mathrm{IC}_1(\boldsymbol{\beta}, \vartheta)$ such that $E\{\mathrm{IC}_1(\boldsymbol{\beta}_0, \vartheta_0)\} = 0$ and

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = n^{-1}\sum_{i=1}^{n} \mathrm{IC}_i(\boldsymbol{\beta}_0, \boldsymbol{\vartheta}_0) + o(n^{-1/2} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\|). \tag{10}$$

If we can show that

$$\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}} = n^{-1}\sum_{i=1}^{n} (\zeta_i - 1)\mathrm{IC}_i(\boldsymbol{\beta}_0, \boldsymbol{\vartheta}_0) + o(n^{-1/2} + \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| + \|\hat{\boldsymbol{\vartheta}}^* - \boldsymbol{\vartheta}_0\|),$$

(11)

then we conclude that, conditional on the observed data, $n^{1/2}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$ has the same asymptotic distribution as $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Regardless of the tools one employs, the basic goal is to establish (10)–(11); see also, Kosorok (2008, Theorem 10.4). This point guides our investigation below.

3.1 Statistical Inference for $\hat{\boldsymbol{\beta}}$

In our resampling scheme, we consider the effect of estimating first the nuisance parameters that are used to calibrate and define the estimating function of interest $\mathbb{S}_{\beta}(\boldsymbol{\theta}_0)$. Here, we proceed along the lines of Jin et al. (2006), who proposed a two-stage least-squares-type coefficient estimator in the accelerated lifetime model.

Let $\mathcal{F}_{obs} = \sigma\{(\delta_i Y_i, X_i, \delta_i, \mathbf{z}_i), i = 1,\ldots, n\}$, $e_i(\boldsymbol{\vartheta}) = \log X_i - \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\vartheta}$, and define the perturbed Gehan estimator for the time-scale parameters,

$$\hat{\boldsymbol{\vartheta}}^* = \arg\min_{\boldsymbol{\vartheta}\in\Re^d} \sum_{i=1}^{n}\sum_{j=1}^{n} \zeta_i\zeta_j\delta_i\{e_j(\boldsymbol{\vartheta}) - e_i(\boldsymbol{\vartheta})\}I\{e_i(\boldsymbol{\vartheta}) - e_j(\boldsymbol{\vartheta}) \le 0\}. \tag{12}$$

Subsequently, define the perturbed mark-scale Gehan-type coefficient estimate $\hat{\boldsymbol{\beta}}^*$ as the solution to the system of equations, $0 = \mathbb{S}^*_{\beta}(\boldsymbol{\beta}, \hat{\boldsymbol{\vartheta}}^*)$, with $\gamma^*(t, \boldsymbol{\vartheta}) = n^{-1}\sum_{i=1}^{n} \zeta_i R_i(t, \boldsymbol{\vartheta})$. Recall, when $\psi(y) = y$, the estimator $\hat{\boldsymbol{\beta}}^*$ is exactly a solution to a perturbed linear system, i.e.

$$\hat{\boldsymbol{\beta}}^* = \left[\sum_{i=1}^{n} \int_{-\infty}^{\infty} \zeta_i\gamma^*(u, \hat{\boldsymbol{\vartheta}}^*)\{\mathbf{z}_i - \bar{\mathbf{z}}^*(u, \hat{\boldsymbol{\vartheta}}^*)\}\mathbf{z}_i^{\mathrm{T}}\mathrm{dN}_i(u, \hat{\boldsymbol{\vartheta}}^*)\right]^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \zeta_i\gamma^*(u, \hat{\boldsymbol{\vartheta}}^*)\{\mathbf{z}_i - \bar{\mathbf{z}}^*(u, \hat{\boldsymbol{\vartheta}}^*)\}Y_i\mathrm{dN}_i(u, \hat{\boldsymbol{\vartheta}}^*),$$

$\bar{\mathbf{z}}^*(u, \boldsymbol{\vartheta}) = \sum_{j=1}^{n} \zeta_j\mathbf{z}_j R_j(u, \boldsymbol{\vartheta}) / \sum_{j=1}^{n} \zeta_j R_j(u, \boldsymbol{\vartheta})$ and $\boldsymbol{\gamma}^*(u, \boldsymbol{\vartheta})$ was defined earlier. Using arguments similar to Jin et al. (2006), one can show that conditional on $\mathcal{F}_{obs}$, $n^{1/2}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}_0)$ has the same asymptotic distribution as $n^{1/2}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)$. Then, a confidence interval is

constructed by repeatedly generating ($\zeta_1, \ldots, \zeta_n$) and solving for $\hat{\boldsymbol{\beta}}^*$ a large number of times, then using the percentiles from the empirical distribution of resampled estimates.

Analogously, if ($\hat{\boldsymbol{\theta}}_1^*, \ldots, \hat{\boldsymbol{\theta}}_B^*$) are the perturbed coefficient vectors obtained from perturbing the data B times, then $B^{-1}\sum_{b=1}^{B}(\hat{\boldsymbol{\theta}}_b^* - \hat{\boldsymbol{\theta}}_0)^{\otimes 2}$ is an estimator for Σ/n. Small sample studies to evaluate the resampling procedures for variance estimation are given in Web Appendix B.

3.2 Statistical Inference for $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}}$

Similarly, let $\hat{\boldsymbol{\beta}}_\lambda^*$ be a root to the perturbed penalized estimating function $\overset{*}{\underset{\beta,\lambda}{}}(\boldsymbol{\beta}, \hat{\boldsymbol{\vartheta}}^*)$, where $\hat{\boldsymbol{\vartheta}}^*$ is the perturbed time-scale coefficient estimate, e.g. see (12) for Gehan weight, and

$$\overset{*}{\underset{\beta,\lambda}{}}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = \overset{*}{\underset{\beta}{}}(\boldsymbol{\beta}, \boldsymbol{\vartheta}) - n\mathbf{q}_\lambda(|\boldsymbol{\beta}|)\mathrm{sgn}(\boldsymbol{\beta}).$$

However, when $\psi(y) = y$, $\overset{*}{\underset{\beta,\lambda}{}}(\boldsymbol{\beta}, \vartheta_0)$ is the quasi-gradient of

$$\overset{*}{\underset{\beta,\lambda}{}}(\boldsymbol{\beta}, \vartheta_0)\frac{1}{2}\{\mathbf{V}^*(\vartheta_0) - \mathbf{A}^*(\vartheta_0)\boldsymbol{\beta}\}^{\mathrm{T}}\{\mathbf{V}^*(\vartheta_0) - \mathbf{A}^*(\vartheta_0)\boldsymbol{\beta}\} + n\sum_{j=1}^{d} p_{\lambda,j}(\beta_j),$$

$\mathbf{V}^*(\boldsymbol{\vartheta}) = \{\mathbf{A}^{*\mathrm{T}}(\boldsymbol{\vartheta})\}^{-1}\mathbf{w}^*(\boldsymbol{\vartheta})$, $\mathbf{A}^*(\boldsymbol{\vartheta})$ is the Choleski decomposition of $\mathbf{M}^*(\boldsymbol{\vartheta})$, and

$$\mathbf{w}^*(\boldsymbol{\vartheta}) = \sum_{i=1}^{n}\zeta_i Y_i \int_{-\infty}^{\infty}\gamma(u, \boldsymbol{\vartheta})\{\mathbf{z}_i - \overline{\mathbf{z}}^*(u, \boldsymbol{\vartheta})\}\mathrm{dN}_i(u, \boldsymbol{\vartheta}),$$
$$\mathbf{M}^*(\boldsymbol{\vartheta}) = \sum_{i=1}^{n}\zeta_i \int_{-\infty}^{\infty}\gamma(u, \boldsymbol{\vartheta})\{\mathbf{z}_i - \overline{\mathbf{z}}^*(u, \boldsymbol{\vartheta})\}\mathbf{z}_i^{\mathrm{T}}\mathrm{dN}_i(u, \boldsymbol{\vartheta}).$$

Therefore, $\hat{\boldsymbol{\beta}}_{\beta,\lambda}^* = \arg\min_{\boldsymbol{\beta}\in\Re^d}\overset{*}{\underset{\beta,\lambda}{}}(\boldsymbol{\beta}, \hat{\boldsymbol{\vartheta}}^*)$. Let $\mathscr{A} = \{j|\beta_{0j} \neq 0\}$, $\boldsymbol{\beta}_{\mathscr{A}} = \{\beta_{0j}|j \in \mathscr{A}\}$, $\hat{\boldsymbol{\beta}}_{\mathscr{A}} = \{\hat{\boldsymbol{\beta}}_{\lambda,j}|j \in \mathscr{A}\}$, and $\hat{\boldsymbol{\beta}}_{\mathscr{A}}^* = \{\hat{\beta}_{\lambda,j}^*|j \in \mathscr{A}\}$. By adapting the arguments of Minnier et al. (2011) to the substitution estimator here, under the conditions given in Web Appendix A, one can show that conditional on $\mathscr{F}_{obs}$, $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}^* - \hat{\boldsymbol{\beta}}_{\mathscr{A}})$ has the same asymptotic distribution as $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathscr{A}} - \boldsymbol{\beta}_{\mathscr{A}})$. Furthermore, $\lim_n P(\hat{\beta}_{\lambda,j}^* \neq 0, \text{for } j \in \mathscr{A}^c|\mathscr{F}_{\mathrm{obs}}) = 0$. As stated in Remark 3 of Johnson et al. (2008) as well as Zou and Li (2008), conditions A1-A2 apply to many penalty functions including bridge penalty (Frank and Friedman, 1993), hard thresholding (Antoniadis, 1997), scad (Fan and Li, 2001), adaptive lasso (Zou, 2006), and logarithmic penalty (Zou and Li, 2008; Johnson et al., 2008). Thus, at a minimum, the distributional result for $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathscr{A}} - \boldsymbol{\beta}_{\mathscr{A}})$ applies to these regularized coefficient estimators.

To improve the finite sample perform of the variance estimators and for comparison purposes, we consider resampling plans similar to those described in Minnier et al. (2011). We consider four different types of confidence intervals based on resamples of the regularized estimates, three of which were considered in Minnier et al. (2011). Let B be the

user-defined total number of resamples and define $\hat{\pi}_{0j} = \#\{\hat{\beta}_{\lambda,j} = 0\}/B$, the empirical estimate that $\beta_j = 0$ based on the bootstrap resamples for $j = 1, \ldots, d$. For asymptotically normal confidence intervals, $\mathrm{CI}_j^{\mathrm{N}}$, we considered the centered intervals: $\hat{\beta}_{\lambda,j} \pm z_{1-\alpha/2}\hat{\sigma}_j$, $\hat{\sigma}_j^2$ is the bootstrap estimate of the sampling variance, $\hat{\sigma}_j^2 = B^{-1}\sum_{l=1}^{B}(\hat{\beta}^*_{\lambda,j,l} - \hat{\beta}_{\lambda,j})^2$. To be consistent with the procedures in Minnier et al. (2011), we set $\mathrm{CI}_j^{\mathrm{N}} = \{0\}$ if $\hat{\pi}_{0j} \geq \hat{p}_{\mathrm{high}}$, $\hat{p}_{\mathrm{high}}$ is an upper threshold such that $\hat{p}_{\mathrm{high}} \to (1 - \alpha)$. The percentile interval $\mathrm{CI}_j^{\mathrm{P}}$ defines upper and lower endpoints through the $(1 - \alpha/2)$-sample quantiles of the bootstrap resamples.

In addition to normal and percentile method, we also implemented two versions of highest density intervals (Minnier et al., 2011). Write the conditional density of $\hat{\beta}_j^*|\mathscr{F}_{\mathrm{obs}}$ as the mixture density $\hat{\pi}_{0j}I(b{=}0) + (1 - \hat{\pi}_{0j})p_j^*(b)$, where $p_j^*(b)$ is the probability density function of $\hat{\beta}_j^*$ provided that it is non-zero. Then, the confidence interval is defined piecewise as follows:

$$\mathrm{CI}_j^{\mathrm{M}} = \begin{cases} \{0\} & \text{(a).Strong evidence } \beta_j{=}0 \\ \{b|p_j^*(b) \geq \kappa_1\} \cup \{0\} & \text{(b).Weak evidence } \beta_j{=}0 \\ \{b|p_j^*(b) \geq \kappa_2\} \cup \{0\} & \text{(c).Weak evidence } \beta_j \neq 0 \\ \{b|p_j^*(b) \geq \kappa_3\} & \text{(d).Strong evidence } \beta_j \neq 0 \end{cases} \quad (13)$$

where $\mathscr{R}(\kappa) = \int I\{p_j^*(b) \geq \kappa\}p_j^*(b)\mathrm{d}b$, $\mathscr{R}(\kappa_1) = (1 - \alpha - \hat{\pi}_{0j})/(1 - \hat{\pi}_{0j})$, $\mathscr{R}(\kappa_2) = (1 - \alpha) + \alpha(\hat{\pi}_{0j} + \hat{p}_{\mathrm{low}})$, $\mathscr{R}(\kappa_3) = (1 - \alpha)$, $\hat{p}_{\mathrm{low}} \to 0$ and $\hat{p}_{\mathrm{high}} \to (1 - \alpha)$. The rules proposed by Minnier et al. (2011) for determining strong and weak evidence depended on the proportion $\hat{\pi}_{0j}$ and the thresholds, $\hat{p}_{\mathrm{high}}$ and $\hat{p}_{\mathrm{low}}$: (a) $\hat{\pi}_{0j} \geq \hat{p}_{\mathrm{high}}$, (b) $\hat{p}_{\mathrm{low}} \leq \hat{\pi}_{0j} < \hat{p}_{\mathrm{high}}$, (c) $\alpha \leq \hat{\pi}_{0j} < \max(\alpha, \hat{p}_{\mathrm{low}})$, and (d) $\hat{\pi}_{0j} < \alpha$. The four cases in $\mathrm{CI}_j^{\mathrm{M}}$ are initially based on the mixture distribution of $\hat{\beta}_j^*|\mathscr{F}_{\mathrm{obs}}$, but Minnier et al. (2011) inflate or deflate the region of each case to improve finite sample performance of the interval estimator. For example, when the evidence is strong for $\beta_j = 0$ or $\beta_j \neq 0$, one ignores the mass in the continuous $p_j^*(b)$ or the atom at zero, respectively. The interval estimator is inflated when there is weak evidence of a non-zero effect in case (c). The exception is when $\hat{p}_{\mathrm{low}} \leq \hat{\pi}_{0j} < \hat{p}_{\mathrm{high}}$ in case (b) where the interval estimator follows directly from the definition of a highest density region for a mixture distribution.

This leads to a second definition of highest density region that reflects the mixture density $\hat{\pi}_{0j}I(b{=}0) + (1 - \hat{\pi}_{0j})p_j^*(b)$ as is, without artificially manipulating the region for weak effects. Note that case (b) in (13) follows directly from the definition of bivariate mixture for a level-α region while cases (a) and (d) are ϵ-approximations to the mixture density since $\hat{\pi}_{0j}$ is very close to one or zero, respectively. So, case (c) in (13) for weak effects is the only one that deviates substantially from the mixture density. By combining cases (b)–(c) of $\mathrm{CI}_j^{\mathrm{M}}$, we take a more agnostic approach to interval estimation for any effect that lacks strong

evidence as an important or unimportant variable. The simple HDR confidence interval is given by

$$\mathrm{CI}_j^{\mathrm{S}} = \begin{cases} \{0\} & \text{(a).Strong evidence } \beta_j = 0 \\ \{b | p_j^*(b) \geq \kappa_1\} \cup \{0\} & \text{(b).Indeterminate} \\ \{b | p_j^*(b) \geq \kappa_3\} & \text{(c).Strong evidence } \beta_j \neq 0 \end{cases} \quad (14)$$

where case (b) is in effect if $\alpha \leq \hat{\pi}_{0j} < \hat{\rho}_{\text{high}}$. Since the confidence interval $\mathrm{CI}_j^{\mathrm{S}}$ in (14) is defined identically for cases (a), (b), and (d) from $\mathrm{CI}_j^{\mathrm{M}}$ in (13), then we can evaluate directly the effect of artificially inflating the region of weak effects in case (c) of $\mathrm{CI}_j^{\mathrm{M}}$.

4. Analysis of Medical Expenditures in SWOG 9509

The randomized Southwest Oncology Group (SWOG) 9509 trial was designed to investigate Paclitaxel plus Carboplatin versus Vinorelbine plus Cisplatin therapies in untreated patients with advanced nonsmall cell lung cancer. The primary study endpoint was survival time (Kelly et al., 2001) and subsequent secondary analyses considered lifetime medical costs (Huang and Lovato, 2002; Huang, 2002). For each of 408 eligible study participants, the lifetime medical cost endpoint was computed from resource utilization metrics, including medications, medical procedures, different treatments on- and off-protocol, and days spent in the outpatient or inpatient clinic. The cost incurred for each type of resource used was computed using national databases and were standardized to 1998 US dollars (Huang, 2002). Resource utilization was measured at 3, 6, 12, 18, and 24 month clinic visits. Both time and cost are modeled on the natural logarithmic scale.

Our analysis merges the cost data from Huang (2002) with another data set of demographic and clinical variables. In all, we considered 18 baseline variables as main effects in the bivariate accelerated failure time model (3). The regressors are treatment arm (tx, 1=Paclitaxel plus Carboplatin, else 0), gender (sex), progression status (prog.stat.), performance status (ps), clinical stage (stage), IIIB by pleural effusion, weight (kg), height (cm), creatinine (creat), albumin (g/dl), calcium (mg/dl), serum lactate dehydrogenase (ldh, U/l), alkaline phosphatase (alkptase), bilirubin (mg/dl), white blood cell count (wbc, cells/microliter), platelet count (platelet, cells/microliter), hemoglobin level (hgb, g/dl), and age (years). Serum lactate dehydrogenase (ldh), alkaline phosphatase (alkptase), and bilirubin are all derived binary random variables, with one indicating that the patient's measurement exceeded the upper limit of normal (ULN). After missing data was removed, we were left with a final sample size of $n = 343$. Gehan weight function $\gamma(u, \vartheta)$ was used throughout.

Table 1 presents coefficient and standard error estimates for the full model using all 18 covariates. We immediately see that in addition to treatment, there are a number of covariates that are important for one of time or cost but not both. Progression status, performance status, albumin, calcium, bilirubin, and white blood cell count are important for both time and cost. Height, serum lactate dehydrogenase, alkaline phosphatase, and hemoglobin are important for time but not cost while treatment is important for cost but not

time. Sex and age are weak predictors for cost and time, respectively, but are not statistically significant at the nominal level. In subsequent analyses using penalized estimating functions, we standardized each of 18 baseline variables to have mean zero and unit variance. This explains why, in Table 1, we present coefficient and standard error estimates for both standardized and non-standardized covariates.

The results of our model selection procedures using the BIC-type criterion are presented in Table 2. Of the 18 variables, we found that when we regularize the cost-scale estimating function only, 6-7 variables are associated with medical cost, including treatment, sex, progression status, performance score, albumin, white blood cell count, and hemoglobin level. Hemoglobin level is a weak effect and adaptive lasso sets the coefficient estimate exactly to zero. We also performed variable selection using the joint regularization approach and found that the best fit model included 13 or 10 cost-scale predictors for lasso or adaptive lasso, respectively. Hence, the models found via joint regularization are approximately 40-50% more complex in the cost-scale than models found via cost-scale regularization only. This point is further explored in Section 5 through simulation studies.

The results of the inference procedures are displayed in the columns of $\hat{\pi}_{0j}$s in Table 2 and in Figure 1. Progression status was the only predictor among the 18 in our data set that was a very strong predictor of lifetime medical cost. Of 1000 resampled data sets, neither the adaptive lasso nor lasso set the coefficient estimate for progression status to zero for any resampled data set. On the other hand, treatment, sex, performance score, albumin, white blood cell count, and hemoglobin level were weak effects. For each coefficient, Figure 1 displays the 95% confidence intervals using normal (dashed), percentile (light gray), and highest density methods (black). Progression status is the only strong effect and therefore the only HDR interval that does not contain the singleton {0}. Note that for weak effects, the normal confidence intervals cross zero whereas the percentile method and HDR intervals do not. Adaptive lasso coefficient estimates for sex, performance score, albumin, white blood cell count, and hemoglobin level are extremely close to zero and the confidence intervals, or lack thereof, reflect this fact. Finally, because there was only modest differences between two versions of HDR confidence interval, only HDR(M) from Section 3 was presented in Figure 1.

5. Simulation Studies

5.1 Comparison of Point Estimators

In the first numerical example, we compare variable selection procedures that regularize in the cost-scale only versus simultaneous regularization of time- and cost-scale parameters altogether. As part of this numerical study, we also evaluate the sensitivity of $\hat{\boldsymbol{\beta}}_\lambda$ to the choice of calibration vector $\boldsymbol{\vartheta}$. The results here can be used to assess the potential loss in efficiency and predictive ability of $\hat{\boldsymbol{\beta}}_\lambda$ when using the time-scale coefficient estimate derived from the full model.

We start by simulating $d = 8$ predictors as standard normal random variables $\mathbf{z} = (z^{(1)}, \ldots, z^{(8)})$ such that $\text{corr}(z^{(j)}, z^{(k)}) = 0.5^{|k-j|}$, for $j, k = 1, \ldots, 8$. Then, we simulated bivariate normal errors, where $\text{var}(\varepsilon_Y) = \text{var}(\varepsilon_T) = \sigma^2$ and $\text{corr}(\varepsilon_Y, \varepsilon_T) = 0.5$ and subsequently defined

(Y, T) according the joint model in (3). The non-zero time-scale coefficients are $\vartheta_1 = \vartheta_5 = \vartheta_8 = 1$, the non-zero cost-scale coefficients are $\beta_1 = 3$, $\beta_2 = 3/2$, and $\beta_5 = 1$ while all other time- and cost-scale coefficients are zero. Note, just as in the SWOG data example, some predictors are important for time only, some are important for cost only, while other predictors affect both cost and time. The censoring random variable C was uniformly distributed Un(0, 6) and the observed data defined accordingly. The data generating process was repeated independently for n observations.

We evaluate the procedures using three summary statistics: median model error, false positive and false negative rates. The cost-scale model error is defined ME = $(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}_0)^{\mathrm{T}} E(\mathbf{z}\mathbf{z}^{\mathrm{T}})(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}_0)$, the false negative rate is the average number of non-zero coefficients incorrectly set to zero, and the false positive rate is average number of coefficients whose true value is zero but estimate is non-zero. We refer to Huang's (2002) coefficient estimators as unregularized estimators and the oracles use regressors from the true subset of non-zero regression coefficients in time- and then cost-scale. In addition to computing the estimators described in Section 2, we also compute two similar but hypothetical estimators. The two hypothetical estimators are computed in exactly the same way as $\hat{\boldsymbol{\beta}}_\lambda$ except that they calibrate with the true time-scale coefficient $\boldsymbol{\vartheta}_0$ or oracle estimator $\boldsymbol{\vartheta}_{\mathscr{A}}$. Adaptive lasso penalty is used for all regularized estimators and tuning via BIC-type criterion. To gauge the success of the combined estimator $\tilde{\boldsymbol{\theta}}_\lambda$ in selecting important time-scale variables, we computed the regularized coefficient estimator in the accelerated lifetime model (Johnson, 2009). Table 3 presents simulation results over 100 Monte Carlo datasets.

The first observation is that calibrating $\mathbb{S}_{\beta,\lambda}(\boldsymbol{\theta})$ in (6) with the full model coefficient estimator $\hat{\boldsymbol{\vartheta}}_0$ is an effective variable selection strategy if evaluated with cost-scale metrics. If we calibrated with the true value $\boldsymbol{\vartheta}_0$ or oracle estimator $\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$, the cost-scale model error, false positive and negative rates are so similar that they are within error of the Monte Carlo study. Even for $n = 90$ with modest censoring, the estimator $\hat{\boldsymbol{\beta}}_\lambda$ possesses operating characteristics close to the oracle $\hat{\boldsymbol{\beta}}_{\mathscr{A}}$. Our second observation is that joint regularization reduces model error and false positive rate in both time- and cost-scale. But, to accomplish this task, the joint variable selection procedure balances good performance across both scales and, hence, the cost-scale performance of $\tilde{\boldsymbol{\beta}}_\lambda$ is somewhat less than that of $\hat{\boldsymbol{\beta}}_\lambda$ in Table 3. For example, the median model error of $\tilde{\boldsymbol{\beta}}_\lambda$ is about twice that of $\hat{\boldsymbol{\beta}}_\lambda$ at $n = 90$ and $\sigma = 1$. At the same time, $\tilde{\boldsymbol{\vartheta}}_\lambda$ reduces model complexity and model error in the time-scale which is obviously not the case with the unregularized time-scale coefficient estimator $\hat{\boldsymbol{\vartheta}}_0$.

5.2 Comparison of Interval Estimators

To examine the operating characteristics of different interval estimators, we conducted numerous simulation studies in the context of our application. We simulated data according to the bivariate accelerated failure time model in (3), with standard normal regressors that followed a first-order Markov model $\mathrm{corr}(z^{(j)}, z^{(k)}) = (0.5)^{|j-k|}$ and errors that followed a bivariate normal distribution with mean zero, unit variance and covariance $\mathrm{cov}(\varepsilon_Y, \varepsilon_T) = 0.5$. The time-scale regression coefficients were all equal to one whereas the mark-scale model is chosen to have varying levels of complexity (Tibshirani and Knight, 1999). For each of four

different designs, the true regression coefficients are clustered in two groups as described in the following two steps.

1. Set the initial coefficients to $\beta_{7+k,h} = \beta_{14+k,h} = (h - k)^2$, $|k| < h$, $h = 1,\ldots, 4$;
2. Scale the initial coefficient values to yield a theoretical $R^2 = 0.75$, where we define $R^2 = \{\beta^{\mathrm{T}} E(\mathbf{z}_1 \mathbf{z}_1^{\mathrm{T}})\beta\} / \{\beta^{\mathrm{T}} E(\mathbf{z}_1 \mathbf{z}_1^{\mathrm{T}})\beta + E(\varepsilon_Y^2)\}$.

Censoring times were independently generated from Un(0, 6) distribution. For all scenarios, observed data are generated for a random sample of size $n = 75$. The simulation scenarios are designed after numerical studies proposed by earlier authors (Tibshirani and Knight, 1999; Wu et al., 2007; Johnson, 2008), however, none of these earlier works focused on interval estimation and none considered induced censoring.

In the current simulation studies, we considered the four interval estimators discussed earlier in Section 3 for several penalty functions discussed in the literature. These penalty functions include bridge (Frank and Friedman, 1993), lasso (Tibshirani, 1996), hard thresholding (Antoniadis, 1997), scad (Fan and Li, 2001), adaptive lasso (Zou, 2006), and logarithmic penalty (Zou and Li, 2008; Johnson et al., 2008). Due to space limitations, part of the simulation results are given in Table 4 while most results have been moved to Supplementary Material. As in Minnier et al. (2011), we used the `hdrcde` package in `R` to implement the highest density regions. Here, we used an Epanechnikov kernel for density estimation with Silverman's rule-of-thumb bandwidth. Otherwise, default settings of the `hdr()` function were used. Although Minnier et al. (2011) proposed thresholds for adaptive lasso based on the Gaussian linear model with orthonormal design, it is not clear the same definitions should apply uniformly across the class of penalty functions considered here. For this reason, we adopted static cutoffs of $\hat{p}_{\text{high}} \equiv p_{\text{high}} = 0.95$ and $\hat{p}_{\text{low}} = 0.49$, the upper bound in the definition proposed by Minnier et al. (2011, p. 1374). Note, that our static definition of $\hat{p}_{\text{low}}$ does not satisfy $\hat{p}_{\text{low}} \to 0$ and so the interval $\alpha \le \hat{\pi}_{0j} < \max(\alpha, \hat{p}_{\text{low}})$ that defines weak evidence of $\beta_j \neq 0$ in $\mathrm{CI}_j^{\mathrm{M}}$ may be too liberal. However, our definition of the simple interval estimator $\mathrm{CI}_j^{\mathrm{S}}$ effectively sets $\hat{p}_{\text{low}} = \alpha = 0.05$ and allows for an indirect investigation into the effects of our proposed static rule for $\hat{p}_{\text{low}}$ in $\mathrm{CI}_j^{\mathrm{M}}$.

In Table 4, we present the empirical coverage probabilities (ECP) and interval lengths of 95% confidence intervals averaged over coefficients belonging to the active and inactive sets. Due to the construction of the simulation scenarios, the cardinality of the active set ranges from 2 to 14 in Models 1 and 4, respectively, and the cardinality of the inactive set is defined as the difference from $d = 21$. Results for all methods are based on 1000 resamples for each of 100 Monte Carlo data sets. The average length of the confidence intervals in the inactive set was longest for HDR(M) and shortest for HDR(S). In the active set, the normal-type interval estimator had the ECP close to the nominal coverage across all four models, the ECP of HDR(M) was similar to the percentile method, and HDR(S) had the worst coverage among the four intervals estimators considered. Compared to the percentile method with similar coverage, the HDR(M) intervals were longer in Models 2–4 and shorter in Model 1. In general, we found that the simple interval estimator HDR(S) had shorter intervals than HDR(M) but also substantially worse coverage probabilities. These observations led us to

conclude that inflating the confidence region for weak effects as proposed by Minnier et al. (2011) results in measurable improvement over a literal interpretation of the mixture distribution of $\hat{\beta}_j^* | \mathscr{F}_{\text{obs}}$ to guide the HDR interval estimator as in CI_j^{S}. Finally, although the resampling methods are only a heuristic for the lasso, we found that, on the active set, the normal and HDR(M) confidence intervals cover the true value close to the nominal level for Models 2–4, on average, while the percentile method under covers by about 5 percentage points.

6. Discussion

In this paper, we proposed model selection and inference methods for censored lifetime medical cost by applying a theory of penalized estimating function to Huang's (2002) calibration estimator. We regarded $\boldsymbol{\beta}$ and $\boldsymbol{\vartheta}$ as the primary and secondary parts, respectively, of a statistical model (Boos and Stefanski, 2013, 1.2) and, as such, estimated $\boldsymbol{\vartheta}$ as simply as possible that led to an estimator $\hat{\boldsymbol{\beta}}_\lambda$ with the desired asymptotic properties. At the request of two anonymous reviewers, we also investigated joint regularized estimation of time- and cost-scale parameters together using penalized estimating functions via least-squares approximation (LSA). In theory, both $\hat{\boldsymbol{\beta}}_\lambda$ and $\tilde{\boldsymbol{\beta}}_\lambda$ will possess the oracle property but our simulation studies suggest that $\hat{\boldsymbol{\beta}}_\lambda$ is superior to $\tilde{\boldsymbol{\beta}}_\lambda$ in finite samples if evaluated with cost-scale metrics as in Table 3. But if we consider the combined model error and complexity across both cost- and time-scales, then joint regularization is better. A third alternative to what was presented in Section 2 is to consider a richer class of methods that has separate regularizations on the cost- and time-scale. Tuning such a procedure could be accomplished using $D(\boldsymbol{\theta})$ defined in Section 2.3 and information criteria defined analogously. But, in this case, the optimal regularization parameters would be computed by minimizing the information criteria over a 2-dimensional surface of regularization parameters rather than over a line, as we proposed via LSA. Thus, although the point estimate from two separate regularizations in time and cost may not be expensive to compute, tuning such a procedure would be significantly more complex computationally than our joint regularization method and post-selection inference would consequently be more laborious as well.

Finally, we investigated post-selection inference procedures by Minnier et al. (2011) applied to our regularized calibration estimator. Although the penalized calibration regression estimator does not, strictly speaking, belong to the class of estimators considered in Minnier et al. (2011), the post-selection inference technique operates exactly the same way after the resampling plan is justified. For adaptive lasso, scad, and similar penalty functions, we found that confidence intervals based on the normal approximation performed surprisingly well. Confidence intervals based on percentiles and HDR performed well when the signal-to-noise ratio was high and not as well when the signal-to-noise ratio was low. Based on the construction of our simulation studies, this conclusion can be read another way. When the signal is composed of many weak predictors as opposed to a few strong predictors, the coverage probability of percentile and HDR confidence intervals on the active set decreases below the nominal level. As stated in Minnier et al. (2011), limitations of HDR can partly be attributed to difficulty in accurately identifying the active set $\mathscr{A}$ in finite samples. At the same time, we found that HDR performed well in finite samples when there are small

number of strong predictors. The real difficulty arises when there are many weak effects, and this may be a deficiency of the coefficient estimator rather than HDR itself (Pötscher and Schneider, 2009, 2010).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge two anonymous referees and the associate editor whose comments enhanced the paper. We are grateful to patients enrolled in the Southwest Oncology Group Study 9509 and Dr. John Crowley for permission to use the data. Dr. Johnson was supported in part by the University of Rochester Center for AIDS Research grant P30AI078498 (NIH/NIAID) and the University of Rochester School of Medicine and Dentistry Clinical & Translational Science Institute. Drs. Long and Johnson were supported in part by a PCORI award (ME-1303-5840) and an NIH/NINDS grant (R21-NS091630) while Dr. Huang was supported in part by grants NIH HL113451 and AI050409. The content is solely the responsibility of the authors and does not necessarily represent the official views of the PCORI or the NIH.

References

Andersen, PK., Borgan, O., Gill, RD., Keiding, N. Statistical models based on counting processes. Springer; New York: 1993.

Antoniadis A. Wavelets in statistics: A review (with discussion). Journal of the Italian Statistical Association. 1997; 6:97–144.

Bang H, Tsiatis AA. Estimating medical costs with censored data. Biometrika. 2000; 87:329–343.

Boos, DD., Stefanski, LA. Essential Statistical Inference. Springer; New York: 2013.

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

Frank IE, Friedman JH. A statistical view of some chemometrics regression tools. Technometrics. 1993; 35:109–148.

Fygenson M, Ritov Y. Monotone estimating equations for censored data. The Annals of Statistics. 1994; 22:732–746.

Huang Y. Calibration regression of censored lifetime medical cost. Journal of the American Statistical Association. 2002; 97:318–327.

Huang Y. Cost analysis with censored data. Medical Care. 2009; 47:S115–S119. [PubMed: 19536024]

Huang Y, Louis TA. Nonparametric estimation of the joint distribution of survival time and mark variables. Biometrika. 1998; 85:785–798.

Huang Y, Lovato L. Tests for lifetime utility or cost via calibrating survial time. Statistica Sinica. 2002; 12:707–723.

Jain AK, Strawderman RL. Flexible hazard regression modeling for medical cost data. Biostatistics. 2002; 3:101–118. [PubMed: 12933627]

Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. Biometrika. 2003; 90(2):341–353.

Jin Z, Lin DY, Ying Z. On least squares regression with censored data. Biometrika. 2006; 93:147–161.

Johnson BA. Variable selection in semiparametric linear regression with censored data. J R Statist Soc Ser B. 2008; 70:351–370.

Johnson BA. Rank-based estimation in the ℓ_1-regularized partly linear model model with application to integrated analyses of clinical predictors and gene expression data. Biostatistics. 2009; 10:659–666. [PubMed: 19553356]

Johnson BA, Lin DY, Zeng D. Penalized estimating functions and variable seleciton in semiparametric regression models. Journal of the American Statistical Association. 2008; 103:672–680. [PubMed: 20376193]

Johnson, BA., Long, Q., Huang, Y., Chansky, K., Redman, M. Log-penalized least squares, iteratively reweighted lasso, and variable selection for censored lifetime medical cost. Department of Biostatistics and Bioinformatics, Emory University; 2012. Technical Report, 2012-02

Kalbfleisch, JD., Prentice, RL. The statistical analysis of failure time data. 2. Wiley; New York: 2002.

Kelly K, Crowley J, Bunn PA Jr, Presant CA, Grevstad PK, Moinpour CM, Ramsey SD, Wozniak AJ, Weiss GR, Moore DR, Israel VK, Livingston RB, Gandra DR. Randomized phase iii trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non-small cell lung cancer: A southwest oncology group trial. Journal of Clinical Oncology. 2001; 19:3210–3218. [PubMed: 11432888]

Kosorok, MR. Introduction to Empirical Processes and Semiparametric Inference. Springer; New York: 2008.

Lin DY. Linear regression analysis of censored medical costs. Biostatistics. 2000; 1:35–47. [PubMed: 12933524]

Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up. Biometrics. 1997; 53:419–434. [PubMed: 9192444]

Minnier J, Tian L, Cai T. A perturbation method for inference on regularized regression estimates. Journal of the American Statistical Association. 2011; 106:1371–1382. [PubMed: 22844171]

Pötscher BM, Schneider U. On the distribution of of the adaptive lasso estimator. Statistical Planning and Inference. 2009; 139:2775–2790.

Pötscher BM, Schneider U. Confidence sets based on penalized maximum likelihood estimators in gaussian regression. Electronic Journal of Statistics. 2010; 4:334–360.

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996; 58(1):267–288.

Tibshirani RJ, Knight K. The covariance inflation factor for adaptive model selection. Journal of the Royal Statistical Society, Series B. 1999; 61:529–546.

Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. The Annals of Statistics. 1990; 18(1):354–372.

Tsiatis, AA. Semiparametric theory and missing data. Springer Verlag; New York: 2006.

Wang H, Leng C. Unified lasso estimation by least squares approximation. Journal of the American Statistical Association. 2007; 102:1039–1048.

Wei LJ, Ying Z, Lin DY. Linear regression analysis of censored survival data based on rank tests. Biometrika. 1990; 77:845–851.

Wu Y, Boos DD, Stefanski LA. Controlling variable selection by the addition of pseudovariables. Journal of the American Statistical Association. 2007; 102:235–243.

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

Zhao H, Tsiatis AA. A consistent estimator for the distribution of quality adjusted survival time. Biometrika. 1997; 84:339–348.

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

Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics. 2008; 36:1509–1533. [PubMed: 19823597]

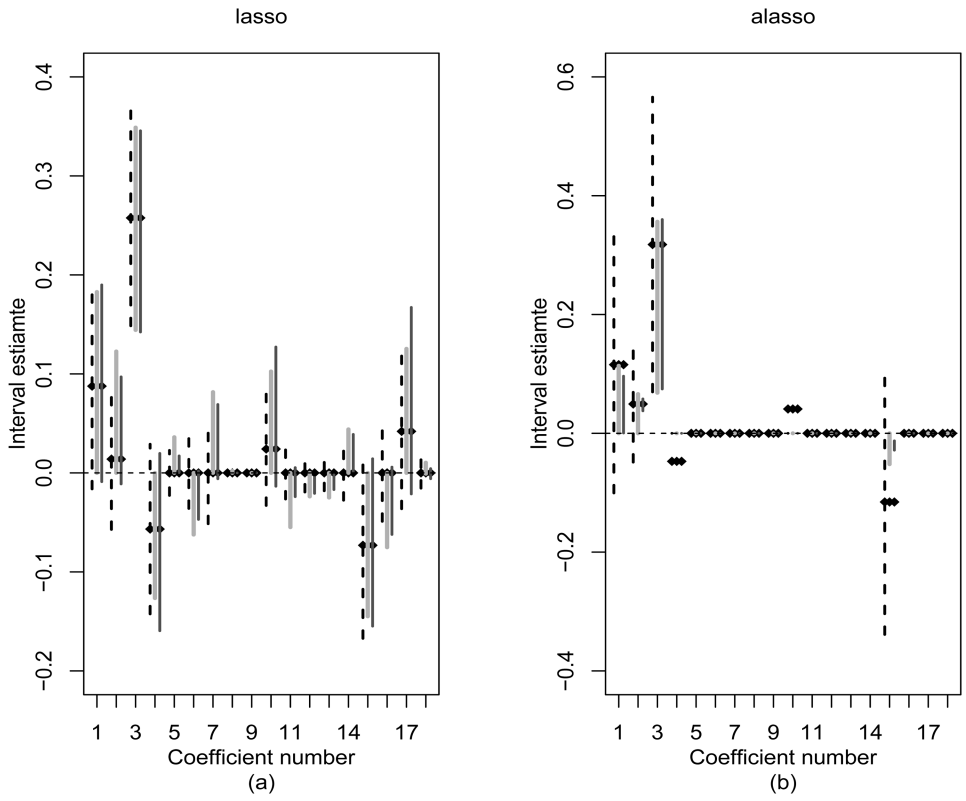

Figure 1.

95% confidence intervals for the penalized cost-scale regression coefficient estimates. Each coefficient estimate has three interval estimates based on methods from Section 3: normal (dashed), percentile (light gray), and HDR (dark gray). Coefficient numbers across the *x*-axis are annotated in Tables 1–2.

Table 1

Mark- and time-scale coefficient and standard error estimates from full model fit to 18 covariates from SWOG 9509. All regularized estimation procedures standardize predictors to have mean zero and unit variance. All table entries are multiplied by 100.

	Standardized X		Unstandardized X	
Name	**Time**	**Cost**	**Time**	**Cost**
1. tx	-2.28 (6.09)	16.98 (4.91)	-4.56 (12.16)	33.68 (9.85)
2. sex	3.79 (8.98)	14.73 (9.21)	8.21 (19.47)	31.94 (20.14)
3. prog.stat	45.93 (11.24)	35.13 (5.15)	157.69 (38.57)	123.55 (18.52)
4. ps	-23.01 (5.58)	-9.67 (5.00)	-47.73 (11.57)	-20.38 (10.45)
5. stage	-11.16 (10.82)	7.81 (9.54)	-34.72 (33.68)	25.63 (29.85)
6. pleural	-3.39 (11.49)	2.59 (9.66)	-13.80 (46.84)	10.54 (39.74)
7. weight	4.67 (6.76)	5.18 (5.88)	0.28 (0.41)	0.33 (0.36)
8. height	-17.42 (8.60)	-10.72 (8.12)	-1.70 (0.84)	-1.06 (0.80)
9. creat	6.30 (13.62)	4.37 (8.50)	13.12 (28.34)	9.14 (17.36)
10. albumin	17.53 (6.89)	12.71 (4.94)	22.15 (8.70)	16.06 (6.28)
11. calcium	-15.02 (6.32)	-9.84 (5.37)	-17.54 (7.37)	-11.26 (6.22)
12. ldh	-21.62 (5.54)	-3.74 (4.86)	-44.44 (11.39)	-6.13 (10.06)
13. alkptase	-12.95 (5.71)	-3.96 (5.43)	-28.14 (12.40)	-7.57 (11.89)
14. bilirubin	8.93 (3.76)	8.17 (4.57)	63.05 (26.58)	51.25 (33.10)
15. wbc	-20.60 (6.80)	-14.47 (5.15)	<0.01 (<0.01)	<0.01 (<0.01)
16. platelet	2.88 (6.07)	-2.71 (5.41)	<0.01 (<0.01)	<0.01 (<0.01)
17. hgb	15.26 (5.98)	6.36 (5.57)	8.41 (3.30)	3.43 (3.06)
18. age	-9.48 (5.97)	-1.32 (5.17)	-1.02 (0.64)	-0.10 (0.57)

Table 2

Estimate penalized regression coefficients for lifetime medical cost in SWOG 9509. Tables columns are coefficient estimates (Est.) and proportion of resampled estimates set to zero ($\hat{\pi}_{0j}$).

	Regularized estimation					
	in cost-only				in cost & time	
	lasso		alasso		lasso	alasso
Name	Est.	$\hat{\pi}_{0j}$	Est.	$\hat{\pi}_{0j}$	Est.	Est.
1. tx	8.76	0.10	11.56	0.86	14.09	14.59
2. sex	1.39	0.59	4.93	0.94	5.02	6.50
3. prog.stat.	25.76	0	31.78	0	31.19	34.07
4. ps	-5.67	0.26	-4.72	0.98	-6.80	-6.78
5. stage	0	0.92	0	1	3.77	2.92
6. pleural	0	0.85	0	1	0	0
7. weight	0	0.67	0	1	2.39	0.13
8. height	0	0.97	0	1	0	0
9. creat	0	0.97	0	1	0	0
10. albumin	2.41	0.34	4.10	0.98	8.60	8.85
11. calcium	0	0.89	0	1	-4.17	-6.16
12. ldh	0	0.94	0	1	-0.42	0
13. alkptase	0	0.93	0	1	0	0
14. bilirubin	0	0.83	0	1	1.99	0
15. wbc	-7.31	0.15	-11.56	0.94	-8.76	-12.03
16. platelet	0	0.66	0	1	-6.28	0
17. hgb	4.20	0.34	0	0.98	5.39	5.19
18. age	0	0.93	0	1	0	0

Table 3

Simulation results comparing model selection procedures. Table entries include the median model error (ME), false positive (FP) rate, and false negative (FN) rate. Cost-scale coefficient estimation is a function of the calibration coefficient vector.

n	σ	Regression & Variable Selection Procedures	Calibration Coefficients	Time-scale Estimates			Cost-scale Estimates		
				ME	FP	FN	ME	FP	FN
60	3	Unregularized	$\hat{\boldsymbol{\vartheta}}_0$	1.65	1.00	0	2.39	1.00	0
		Oracles	$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.48	0	0	0.64	0	0
		Time-scale only	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	1.47	0.17	0.34	-	-	-
		Cost- & time-scale	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	1.33	0.31	0.22	1.80	0.38	0.04
		Cost-scale only	$\boldsymbol{\vartheta}_0$	0	0	0	1.68	0.39	0.04
			$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.48	0	0	1.65	0.38	0.05
			$\hat{\boldsymbol{\vartheta}}_0$	1.65	1.00	0	1.70	0.37	0.04
60	1	Unregularized	$\hat{\boldsymbol{\vartheta}}_0$	0.19	1.00	0	0.22	1.00	0
		Oracles	$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.06	0	0	0.06	0	0
		Time-scale only	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	0.10	0.14	0	-	-	-
		Cost- & time-scale	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	0.10	0.19	0	0.21	0.23	0
		Cost-scale only	$\boldsymbol{\vartheta}_0$	0	0	0	0.14	0.09	0
			$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.06	0	0	0.12	0.11	0
			$\hat{\boldsymbol{\vartheta}}_0$	0.19	1.00	0	0.10	0.10	0
90	1	Unregularized	$\hat{\boldsymbol{\vartheta}}_0$	0.09	1.00	0	0.15	1.00	0
		Oracles	$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.03	0	0	0.05	0.00	0
		Time-scale only	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	0.04	0.07	0	-	-	-
		Cost- & time-scale	$\hat{\boldsymbol{\vartheta}}_{\lambda}$	0.06	0.14	0	0.14	0.19	0
		Cost-scale only	$\boldsymbol{\vartheta}_0$	0	0	0	0.07	0.03	0
			$\hat{\boldsymbol{\vartheta}}_{\mathscr{A}}$	0.03	0	0	0.07	0.04	0
			$\hat{\boldsymbol{\vartheta}}_0$	0.09	1.00	0	0.06	0.04	0

Table 4

Simulation results for interval estimators. All table entries are multiplied by 100.

$\hat{\beta}_\lambda$	Model	Subset	Feature	Normal	Percentile	HDR(M)	HDR(S)
alasso	1	$\mathscr{A}^c$	Coverage	98.1	99.4	98.2	98.2
			Length	65.0	57.1	67.3	54.5
		$\mathscr{A}$	Coverage	94.5	93.0	96.5	92.5
			Length	84.2	76.8	75.4	72.6
	2	$\mathscr{A}^c$	Coverage	97.5	99.2	98.1	98.1
			Length	64.7	57.3	67.9	54.7
		$\mathscr{A}$	Coverage	96.2	90.8	92.2	88.8
			Length	77.7	67.2	75.8	62.8
	3	$\mathscr{A}^c$	Coverage	97.7	99.1	98.0	98.0
			Length	64.1	57.6	68.5	55.0
		$\mathscr{A}$	Coverage	94.0	89.3	89.2	85.7
			Length	73.8	64.4	75.1	60.3
	4	$\mathscr{A}^c$	Coverage	98.1	98.9	98.3	98.3
			Length	63.7	56.8	67.1	53.9
		$\mathscr{A}$	Coverage	94.1	87.8	90.0	84.1
			Length	72.4	63.0	74.6	59.4
lasso	1	$\mathscr{A}^c$	Coverage	97.7	98.1	96.9	96.9
			Length	61.4	54.3	76.8	53.5
		$\mathscr{A}$	Coverage	89.5	90.5	89.5	88.0
			Length	82.1	71.8	71.2	70.0
	2	$\mathscr{A}^c$	Coverage	97.5	98.1	96.6	96.6
			Length	59.7	53.7	75.2	52.8
		$\mathscr{A}$	Coverage	95.7	90.2	94.2	90.2
			Length	69.3	61.3	74.7	59.5
	3	$\mathscr{A}^c$	Coverage	97.5	98.5	96.5	96.5
			Length	59.2	53.6	75.1	52.7
		$\mathscr{A}$	Coverage	94.4	90.1	94.9	89.8
			Length	65.8	58.8	74.9	57.1

$\hat{\beta}_\lambda$	Model	Subset	Feature	Normal	Percentile	HDR(M)	HDR(S)
	4	$\mathscr{A}^c$	Coverage	96.9	98.6	96.9	96.9
			Length	58.7	52.9	74.6	52.1
		$\mathscr{A}$	Coverage	95.0	89.1	94.3	89.9
			Length	64.3	57.5	75.0	55.9