# INFERENCE FOR LOW-DIMENSIONAL COVARIATES IN A HIGH-DIMENSIONAL ACCELERATED FAILURE TIME MODEL

Hao Chai[1], Qingzhao Zhang[2], Jian Huang[3] and Shuangge Ma[1,2]

[1] *Yale University,* [2] *Xiamen University and*
[3] *The Hong Kong Polytechnic University*

*Abstract:* Data with high-dimensional covariates are now commonly encountered. Compared to other types of responses, research on high-dimensional data with censored survival responses is still relatively limited, and most of the existing studies have been focused on estimation and variable selection. In this study, we consider data with a censored survival response, a set of low-dimensional covariates of main interest, and a set of high-dimensional covariates that may also affect survival. The accelerated failure time model is adopted to describe survival. The goal is to conduct inference for the effects of low-dimensional covariates, while properly accounting for the high-dimensional covariates. A penalization-based procedure is developed, and its validity is established under mild and widely adopted conditions. Simulation suggests satisfactory performance of the proposed procedure, and the analysis of two cancer genetic datasets demonstrates its practical applicability.

*Key words and phrases:* AFT model, censored survival data, high-dimensional inference.

## 1. Introduction

Data with high-dimensional covariates but limited sample sizes are now routinely encountered in many fields. In this study, we consider such data with a survival response. With the additional complexity brought by censoring, compared to other types of responses, research on data with a survival response has been relatively limited.

With a sample of $n$ iid observations, consider the AFT (accelerated failure time) model

$$\mathbf{T} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{Z}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}. \tag{1.1}$$

Here $\mathbf{T}$ is a length $n$ vector of event times on the logarithmic scale. The covariate effects contain two components: $\mathbf{X}$ with a fixed $p$ is the $n \times p$ design matrix for

the low-dimensional component, and $\mathbf{Z}$ is the $n \times q$ design matrix for the high-dimensional component, where $q$ can be much larger than $n$. Vectors $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ are regression coefficients, and $\boldsymbol{\epsilon}$ is the vector of random errors with mean zero and variance matrix $\sigma^2 \mathbf{I}$. Assume that the response variable and covariates are properly centered, such that the intercept term is omitted. Our main interest lies in inference (and estimation) for the low-dimensional parameter $\boldsymbol{\beta}_0$. Such an inference problem arises in many applications. As an example, consider a disease treatment study with genetic measurements (Johnson (2009); Morris et al. (2014)). Here the event time is disease prognosis, the low-dimensional covariates may include the treatment variables of main interest, and the high-dimensional covariates may contain genetic markers that contribute to survival but are of secondary interest.

For modeling survival, we adopt the AFT model. Compared to alternative models, for example the Cox model, its advantages include simple interpretation and low computational cost, which are especially desirable with high-dimensional data. Under low-dimensional settings, notable studies on estimation and inference with the AFT model include Buckley and James (1979), Tsiatis (1990), Wei, Ying and Lin (1990), and Stute (1993). Under high-dimensional settings, both variable selection and dimension reduction methods have been developed. In the existing studies, the penalization technique, with its significant theoretical and empirical advantages, has been used by many authors. Examples include Huang, Ma and Xie (2006), Johnson (2008), Cai, Huang and Tian (2009), Huang and Ma (2010), Ma and Du (2012), and Hu and Chai (2013).

Most of the existing studies have been focused on estimation and variable selection. Comparatively, attention on inference is limited. In applications, inference plays an equally important role. For an effect, it is of interest to know not only its level but also its confidence level. Inference with high-dimensional data is a challenging and important problem. Zhang and Zhang (2014), van de Geer et al. (2014), and Javanmard and Montanari (2014) have studied the construction of confidence intervals for low-dimensional parameters in high-dimensional linear and generalized linear models. Bühlmann (2013) considers a corrected ridge regression approach for computing $p$-values for general hypotheses under high-dimensional settings. The focus on the low-dimensional parameters in the present study is similar to that in the aforementioned works. Another relevant study is Voorman, Shojaie and Witten (2014), which constructs a test statistic for each regression coefficient based on the penalized score test. Other approaches

include the least squares after double selection by Belloni, Chernozhukov and Hansen (2014) and the post selection inference by Lee et al. (2016) and Berk et al. (2013). Beyond the linear framework, Yang et al. (2016) studies inference under nonlinear models. Ning and Liu (2017) discusses low-dimensional inference under the general high-dimensional M-estimation framework. For censored data, Fang, Ning and Liu (2016) studies inference for the high-dimensional Cox model. Inference under the high-dimensional AFT model, which provides a flexible alternative to the Cox and other models, is of interest but has not been pursued.

Consider the AFT model (1.1) for right censored survival data. The goal is to conduct inference on the estimate of the low-dimensional parameter $\boldsymbol{\beta}_0$. Different from the existing works on regularized estimation and variable selection, we focus on inference. This is considerably more complicated than the existing inference studies in the context of linear and generalized models due to the presence of censoring. The proposed method has the potential to be extended to other censored survival model settings. In what follows, the proposed method and its statistical properties are established in Sections 2 and 3. Numerical study, including simulation in Section 4 and data analysis in Section 5, is conducted to examine practical performance. The article concludes with discussions in Section 6. Additional technical details are provided in the Supplementary File.

## 2. Inference Under the High-Dimensional AFT Model

Denote $T$ as the logarithm of the event time and $C$ as the logarithm of the censoring time. Let $Y = \min(T, C)$ and $\delta = 1\{T \leq C\}$. Let $X_{p \times 1} = (X_1, X_2, \ldots, X_p)^\top$ and $Z_{q \times 1} = (Z_1, Z_2, \ldots, Z_q)^\top$ be the random vectors of covariables. The observed design matrices $\mathbf{X}$ and $\mathbf{Z}$ are generated from $X$ and $Z$ respectively. For subject $i(= 1, \ldots, n)$, under right censoring, we observe $(y_i, \delta_i, \mathbf{x}_{i,}, \mathbf{z}_{i,})$. Throughout the paper, for a matrix $\mathbf{M}$, let $\mathbf{m}_{i,}$ be the transpose of its $i$th row, $\mathbf{m}_{,j}$ be its $j$th column, and $m_{i,j}$ be its $(i, j)$th element. When $p + q \ll n$ and the column spaces of $\mathbf{X}$ and $\mathbf{Z}$ are of full rank, there are multiple estimation approaches. Here we adopt the weighted least squares approach developed in Stute (1993) and Stute (1996) that has a simpler objective function and is easier to extend to high-dimensional settings than the competing alternatives such as the Bukley-James and rank-based methods.

Consider that $p$ is fixed and small compared to $n$, but $q$ is large (comparable

to or even much larger than $n$). Without loss of generality, assume that the samples are sorted according to the $y_i$'s. The Kaplan-Meier weights are defined as

$$\omega_1 = \frac{\delta_1}{n}, \quad \omega_i = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i-1} \left( \frac{n-j}{n-j+1} \right)^{\delta_j}, \quad i = 2, \ldots, n.$$

Let $\mathbf{W} = \mathrm{diag}\{n\omega_1, n\omega_2, \ldots, n\omega_n\}$. The design matrices and response variable are weighted normalized such that $\mathbf{1}^\top \mathbf{W} \mathbf{x}_{,k} = 0$, $\mathbf{1}^\top \mathbf{W} \mathbf{z}_{,j} = 0$, and $\mathbf{1}^\top \mathbf{W} \mathbf{y} = 0$ for $j = 1, \ldots, q$ and $k = 1, \ldots, p$. Consider the objective function

$$L_0(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2n} \| \mathbf{W}^{1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\theta}) \|_2^2,$$

where $\| \cdot \|_2$ denotes the $\ell_2$-norm. Under mild conditions, Stute (1996) establishes that the estimator obtained by minimizing $L_0(\boldsymbol{\beta}, \boldsymbol{\theta})$ is consistent and asymptotically normal as $n \to \infty$ for fixed $p$ and $q$. Compared to other ways of accommodating censoring, using the Kaplan-Meier weights, as in this approach, is computationally advantageous. The minimizer of $L_0(\boldsymbol{\beta}, \boldsymbol{\theta})$ satisfies the normal equations

$$\begin{cases} \dfrac{\partial L_0(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \dfrac{1}{n} \mathbf{X}^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\theta}) = \mathbf{0}, \\[2mm] \dfrac{\partial L_0(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \dfrac{1}{n} \mathbf{Z}^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\theta}) = \mathbf{0}. \end{cases}$$

This is equivalent to

$$\frac{1}{n}(\mathbf{X} - \mathbf{ZB})^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\theta}) = \mathbf{0},$$

for all $q \times p$ matrices $\mathbf{B}$. The quasi normal equation is

$$(\mathbf{X} - \mathbf{ZB})^\top \mathbf{W} \mathbf{X} \boldsymbol{\beta} = (\mathbf{X} - \mathbf{ZB})^\top \mathbf{W}(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}). \tag{2.1}$$

The estimate of $\boldsymbol{\beta}$ is unbiased, if the estimate of $\boldsymbol{\theta}$ is unbiased, or if $(\mathbf{X} - \mathbf{ZB})^\top \mathbf{W} \mathbf{Z} = \mathbf{0}$, provided that $(\mathbf{X} - \mathbf{ZB})^\top \mathbf{W} \mathbf{X}$ is invertible. Under high-dimensional settings, however, the estimate of $\boldsymbol{\theta}$ is usually biased. Hence it is desirable to find a matrix $\mathbf{B}$ such that $\mathbf{W}^{1/2}(\mathbf{X} - \mathbf{ZB})$ and $\mathbf{W}^{1/2} \mathbf{Z}$ are almost orthogonal, and $(\mathbf{X} - \mathbf{ZB})^\top \mathbf{W} \mathbf{X}$ is invertible. To achieve these two goals, we consider regularized estimation for $\boldsymbol{\theta}$ and $\mathbf{B}$. In this paper, we use LASSO (Tibshirani (1996)) for regularization. Note that other penalties, such as Dantzig (Candes and Tao (2007)), SCAD (Fan and Li (2001)), and MCP (Zhang (2010)) are also applicable. The LASSO penalized estimator is defined as

$$\begin{cases} (\widetilde{\boldsymbol{\beta}}^*, \widetilde{\boldsymbol{\theta}}) = \underset{\boldsymbol{\beta}, \boldsymbol{\theta} \in \mathbb{R}^{p+q}}{\arg\min} \left\{ L_0(\boldsymbol{\beta}, \boldsymbol{\theta}) + \lambda_0 \sum_{j=1}^{q} |\theta_j| \right\}, \\ \widetilde{\mathbf{b}}_{,k} = \underset{\mathbf{b} \in \mathbb{R}^q}{\arg\min} \left\{ L_k(\mathbf{b}) + \lambda_k \sum_{j=1}^{q} |b_j| \right\} \quad \text{for } k = 1, 2, \dots, p. \end{cases} \tag{2.2}$$

Here $L_k(\mathbf{b}) = (1/2n)\|\mathbf{W}^{1/2}(\mathbf{x}_{,k} - \mathbf{Z}\mathbf{b})\|_2^2$, for $k = 1, 2, \dots, p$, and $\lambda_0, \lambda_k > 0$ are data-dependent tuning parameters. In the first objective function in (2.2), penalization is not imposed on $\boldsymbol{\beta}$, as it is low dimensional and our interest lies in conducting inference for its estimate (as opposed to selection). The first estimation in (2.2) generates a good estimate of $\boldsymbol{\theta}_0$. In the second estimation, our goal is to find the aforementioned matrix $\mathbf{B}$. Let $\widetilde{\mathbf{B}} = (\widetilde{\mathbf{b}}_{,1}, \dots, \widetilde{\mathbf{b}}_{,p})$ and $\widetilde{\mathbf{X}} = \mathbf{X} - \mathbf{Z}\widetilde{\mathbf{B}}$. The estimate of $\boldsymbol{\beta}_0$ can be obtained by replacing $\mathbf{B}$ and $\boldsymbol{\theta}$ by $\widetilde{\mathbf{B}}$ and $\widetilde{\boldsymbol{\theta}}$ in (2.1) and satisfies

$$\widetilde{\mathbf{X}}^\top \mathbf{W} \mathbf{X} \widetilde{\boldsymbol{\beta}} = \widetilde{\mathbf{X}}^\top \mathbf{W} \left( \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\theta}} \right).$$

If $\widetilde{\mathbf{X}}^\top \mathbf{W} \mathbf{X}$ is invertible,

$$\widetilde{\boldsymbol{\beta}} = \left( \widetilde{\mathbf{X}}^\top \mathbf{W} \mathbf{X} \right)^{-1} \widetilde{\mathbf{X}}^\top \mathbf{W} \left( \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\theta}} \right). \tag{2.3}$$

In Section , we show that $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal as $n \to \infty$ under mild regularity conditions. The estimate of the asymptotic covariance matrix can be constructed using the observed data (readers can skip the following technical details on $\widetilde{\varphi}$, $\widetilde{\tau}_0$, $\widetilde{\tau}_1^j$, $\widetilde{\tau}_2^j$, and $\widetilde{\psi}_j$ without having trouble understanding the main results). Define

$$\begin{cases} \widetilde{\varphi}_j(\widetilde{\mathbf{x}}_{i,}, y_i, \mathbf{x}_{i,}, \mathbf{z}_{i,}) = \widetilde{x}_{i,j} \left( y_i - \mathbf{x}_{i,}^\top \widetilde{\boldsymbol{\beta}} - \mathbf{z}_{i,}^\top \widetilde{\boldsymbol{\theta}} \right), \\ \widetilde{\tau}_0(y) = \exp \left( \sum_{i:y_i<y,\delta_i=0} \frac{1}{(n - \sum_{k=1}^n 1\{y_k \le y_i\})} \right), \\ \widetilde{\tau}_1^j(y) = \sum_{k:y_k>y,\delta_k=1} \frac{\widetilde{\varphi}_j(\widetilde{\mathbf{x}}_{k,}, y_k, \mathbf{x}_{k,}, \mathbf{z}_{k,})\widetilde{\tau}_0(y_k)}{(n - \sum_{i=1}^n 1\{y_i \le y\})}, \\ \widetilde{\tau}_2^j(y) = \sum_{i:y_i<y,\delta_i=0} \left[ \frac{\{\sum_{k:y_k>y_i,\delta_k=1} \widetilde{\varphi}_j(\widetilde{\mathbf{x}}_{k,}, y_k, \mathbf{x}_{k,}, \mathbf{z}_{k,})\widetilde{\tau}_0(y_k)\}}{(n - \sum_{l=1}^n 1\{y_l \le y_i\})^2} \right]. \end{cases}$$

For the $k$th sample, let

$$\widetilde{\psi}_j(\widetilde{\mathbf{x}}_{k,}, y_k, \mathbf{x}_{k,}, \mathbf{z}_{k,}) = \widetilde{\varphi}_j(\widetilde{\mathbf{x}}_{k,}, y_k, \mathbf{x}_{k,}, \mathbf{z}_{k,})\widetilde{\tau}_0(y_k)\delta_k + \widetilde{\tau}_1^j(y_k)(1 - \delta_k) - \widetilde{\tau}_2^j(y_k).$$

Let $\widetilde{\sigma}_{i,j}$ be the sample correlation of $\widetilde{\psi}_i$ and $\widetilde{\psi}_j$, $\widetilde{\boldsymbol{\Sigma}}_1 = (\widetilde{\sigma}_{i,j})_{p \times p}$, and $\widetilde{\boldsymbol{\Sigma}}_0 = \widetilde{\mathbf{X}}^\top \mathbf{W} \mathbf{X}$. It is not hard to prove that $\widetilde{\boldsymbol{\Sigma}}_0$ and $\widetilde{\boldsymbol{\Sigma}}_1$ satisfy $\widetilde{\boldsymbol{\Sigma}}_0 \xrightarrow{p} \boldsymbol{\Sigma}_0$ and $\widetilde{\boldsymbol{\Sigma}}_1 \xrightarrow{p} \boldsymbol{\Sigma}_1$

using Corollary 1.8 of Stute (1996), where $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are defined in Theorem 1. Hence,

$$\boldsymbol{\Sigma}\left(\widetilde{\boldsymbol{\beta}}\right) = \widetilde{\boldsymbol{\Sigma}}_0^{-1}\widetilde{\boldsymbol{\Sigma}}_1\left(\widetilde{\boldsymbol{\Sigma}}_0^{-1}\right)^{\top} = \left(\widetilde{\mathbf{X}}^{\top}\mathbf{W}\mathbf{X}\right)^{-1}\widetilde{\boldsymbol{\Sigma}}_1\left(\mathbf{X}^{\top}\mathbf{W}\widetilde{\mathbf{X}}\right)^{-1} \tag{2.4}$$

is a consistent estimate of the asymptotic covariance matrix of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. When there is no censoring, $\mathbf{W} = \mathbf{I}$, and the proposed estimate of $\boldsymbol{\beta}_0$ and covariance matrix coincide with those proposed in Zhang and Zhang (2014).

Using Hotelling's $T^2$ statistic, a level $1 - \alpha$ confidence region for $\boldsymbol{\beta}_0$ is

$$D = \left\{\boldsymbol{\beta} \in \mathbb{R}^p : \left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{\top}\widetilde{\boldsymbol{\Sigma}}^{-1}\left(\widetilde{\boldsymbol{\beta}}\right)\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) < \frac{(n-1)p}{(n-p)n}F_{1-\alpha,p,n-p}\right\}.$$

For the $j$th component of $\boldsymbol{\beta}_0$, the marginal confidence interval can be constructed as $\widetilde{\beta}_j \pm t_{1-\alpha/2,n-p} \cdot se(\widetilde{\beta}_j)$, where $se(\widetilde{\beta}_j)$ is the square-root of the $(j,j)$th element of $\widetilde{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{\beta}})/n$.

We conclude this section by providing some heuristics on why $\widetilde{\boldsymbol{\beta}}$ achieves asymptotic normality. Recall that if $\delta_i = 0$, $\omega_i = 0$. Hence we can replace $\mathbf{y}$ in (2.3) by $\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{Z}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}$ and obtain

$$\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}^{\top}\mathbf{W}\mathbf{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}^{\top}\mathbf{W}\boldsymbol{\epsilon} = -\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}^{\top}\mathbf{W}\mathbf{Z}\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

When the term on the right hand side is of a smaller order than the terms on the left, the asymptotic distributions of the two terms on the left hand side are the same. Hence with LASSO penalization, we can effectively reduce the high-dimensional problem to a low-dimensional one.

## 3. Asymptotic Results

We first introduce some notations. Let $\tau_Y$, $\tau_T$, and $\tau_C$ be the end points of the support of $Y, T$, and $C$, respectively. Let $U = \left(X^{\top}, Z^{\top}\right)^{\top}$ and $F$ be the joint distribution of $(U, T)$. Following Stute (1996), we write

$$\widetilde{F}(\mathbf{u}, t) = \begin{cases} F(\mathbf{u}, t), & t < \tau_Y, \\ F(\mathbf{u}, \tau_Y-) + 1\{\tau_Y \in \nu\}F(\mathbf{u}, \{\tau_Y\}), & t \geq \tau_Y, \end{cases} \tag{3.1}$$

where $\nu$ is the set of atoms of $H$, the distribution function of $Y$. Denote $\theta_{0j}$ as the $j$th component of $\boldsymbol{\theta}_0$. Let $A_0 = \{j : \theta_{0j} \neq 0, j = 1, 2, \ldots, q\}$ and $|A_0|$ be the cardinality of $A_0$. There exists a $q \times p$ matrix $\mathbf{B}_0 = (\mathbf{b}_{,1}, \ldots, \mathbf{b}_{,p})$ that satisifes $E_{\widetilde{F}}\{Z(X - \boldsymbol{B}_0^{\top}Z)\} = \mathbf{0}_{q \times p}$. Define

$$\widetilde{H}^{11}(\mathbf{u}, y) = P(U \leq \mathbf{u}, Y \leq y, \delta = 1), \quad \widetilde{H}^0(y) = P(Y \leq y, \delta = 0).$$

For $j = 1, \ldots, p$, let $\varphi_j(U, Y) = \left(X_j - Z^\top \mathbf{b}_{,j}\right)\left(Y - X^\top \boldsymbol{\beta}_0 - Z^\top \boldsymbol{\theta}_0\right)$,

$$\tau_0(y) = \exp\left(\int_0^{y-} \frac{\widetilde{H}^0(ds)}{1 - H(s)}\right),$$

$$\tau_1^j(y) = \frac{1}{1 - H(y)} \int 1\{s > y\} \varphi_j(\mathbf{u}, s) \tau_0(s) \widetilde{H}^{11}(d\mathbf{u}, ds),$$

$$\tau_2^j(y) = \iint \frac{1\{v < y, v < s\} \varphi_j(\mathbf{u}, s) \tau_0(s)}{\{1 - H(s)\}^2} \widetilde{H}^0(dv) \widetilde{H}^{11}(d\mathbf{u}, ds),$$

and $\psi_j = \varphi_j(U, Y) \tau_0(Y) \delta + \tau_1^j(Y)(1 - \delta) - \tau_2^j(Y)$.

**Assumption 1.** $P(T \leq C | T, X, Z) = P(T \leq C | T)$.

**Assumption 2.** *The matrix* $\mathbf{B}_0 = (\mathbf{b}_{,1}, \ldots, \mathbf{b}_{,p})$ *is sparse. If* $K_0 \subset \{1, \ldots, q\}$ *is the index set of the nonzero rows of* $\mathbf{B}_0$, $|K_0| \asymp |A_0|$.

**Assumption 3.** $X_k$, $Z_j$, *and* $\epsilon$ *have sub-Gaussian distributions for* $k = 1, \ldots, p$, $j = 1, \ldots, q$. *Such distributions are from* $c_1(< \infty)$ *distribution families, and each family is determined by at most* $c_2$ *parameters.*

**Assumption 4.** *Denote the indices of columns of* $X$, $Z_{A_0}$, *and* $Z_{K_0}$ *in* $U$ *as* $J_0$, $A_0^+$, *and* $K_0^+$. *Let* $\Gamma = E_{\widetilde{F}}(UU^\top)$ *and* $\|\Gamma\|_\infty = O(1)$. *For* $A = J_0 \cup A_0^+ \cup K_0^+$, *the matrix* $\Gamma$ *satisfies the restricted eigenvalue condition* $RE(|A|)$,

$$\kappa^2(|A|) = \inf_{\|\mathbf{a}\|_{1,A^c} \leq 3\|\mathbf{a}\|_{1,A}} \frac{\mathbf{a}^\top \Gamma \mathbf{a}}{\|\mathbf{a}\|_{2,A}^2} \geq c_* > 0.$$

**Assumption 5.** $\int |\varphi_j(\mathbf{u}, s)| C^{1/2}(s) \widetilde{F}(d\mathbf{u}, ds) < \infty$ *and* $E\{\varphi_j(U, Y) \tau_0(Y) \delta\}^2 < \infty$ *for any* $j = 1, \ldots, p$, *where* $C(s) = \int_0^{s-} G(dy)/[\{1 - H(y)\}\{1 - G(y)\}]$ *and* $G$ *is the distribution function of the censoring variable.*

Under Assumption 1, $\delta$ is conditionally independent of the covariate $U$ given the failure time $Y$. This assumption also specifies that $Y$ and $C$ are independent. However, this assumption does allow the censoring variable to be dependent on the covariates. For more discussions on this assumption, we refer to Stute (1996), Huang, Ma and Xie (2006, 2007). The sparsity Assumption 2 ensures that the LASSO selector converges to the true value at a fast rate. A similar assumption has been made in Fang, Ning and Liu (2016). Under Assumption 3, at most $c_3 \leq c_1 c_2$ parameters are needed to fully determine the distributions of $X_k$, $Z_j$, and $\epsilon$. This assumption has been made in high-dimensional studies and can be weakened at the price of a smaller $q$. Assumptions 4 is standard in the high-dimensional model selection literature. Assumption 5 has been made in Stute (1996) and ensures the asymptotic normality of the proposed estimator.

Denote $\Gamma_{S_1,S_2}$ as the extracted submatrix of $\Gamma$ where $S_1$ is the row index and $S_2$ is the column index. Properties of $\widetilde{\boldsymbol{\beta}}$ can be established as follows.

**Theorem 1.** *Suppose that Assumptions* 1-4 *hold. If* $n \to \infty$, $\sqrt{n}\lambda_{*,p}\lambda_0|A_0| \to 0$, *where* $\lambda_{*,p} = \max\{\lambda_k, 1 \le k \le p\}$, *and with a large enough* $M > 0$, $\min_{k=0,1,\ldots,p}$ $\lambda_k > M\sqrt{\log q/n}$, *then we have*

$$\left\| \frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\boldsymbol{\epsilon} - \frac{1}{\sqrt{n}}(\mathbf{X} - \mathbf{Z}\mathbf{B}_0)^\top \mathbf{W}\mathbf{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right\|_1 \xrightarrow{p} 0.$$

*Together with* Assumption 5, *we have*

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}),$$

*where* $\Sigma_0 = \Gamma_{J_0,J_0} - \Gamma_{K_0^+,J_0}^\top \Gamma_{K_0^+,K_0^+}^{-1} \Gamma_{K_0^+,J_0}$ *and* $\Sigma_1 = (\sigma_{i,j})_{1 \le i,j \le p}$ *with* $\sigma_{i,j} = Cov(\psi_i, \psi_j)$.

The proof is provided in the Supplementary File. With the covariate having two components, two assumptions on the tuning parameters are made. This theorem requires that $\sqrt{n}\lambda_{*,p}\lambda_0|A_0| \to 0$ and $\min_{k=0,1,\ldots,p} \lambda_k > M\sqrt{\log q/n}$. If $\lambda_{*,p} \asymp \lambda_0 \asymp \sqrt{\log q/n}$, the requirement is $\log(q)|A_0|/\sqrt{n} \to 0$ as $n \to \infty$. Thus, for a large $n$, $q = o(\exp(\sqrt{n}))$ if $|A_0|$ is fixed. Compared to the sample size requirement for the selection and estimation error bounds ($q = o(\exp(n))$), a larger sample size is required to obtain asymptotic normality. This requirement is similar to that in Remark 3(a) of Zhang and Zhang (2014) for the simple linear model without censoring.

When defining the proposed method, LASSO is adopted. It has certain computational advantages over concave penalties. Here the LASSO penalty can be replaced by nonconvex penalties such as MCP and SCAD. We conjuncture that the theoretical properties in our theorem still hold, but under slightly different conditions. In numerical studies, we examine MCP along with LASSO.

## 4. Simulation

In simulations, we compared the proposed method with two alternatives. The first is the oracle method, which knows in advance which covariate effects are nonzero and proceeds with a low-dimensional model; this method is only applicable in simulations. The alternative, referred to as "Only.X", analyzes the low-dimensional covariates only. For the penalty used in the proposed method, beyond LASSO, we also applied MCP.

In model (1.1), set $p = 2$, $q = 1,000$, and $|A_0| = 6$. Thus, eight covariates, two from the low-dimensional set and six from the high-dimensional set,

Table 1. Simulation: empirical coverage probability under Case 1. (SS, FR, Cor) = (sample size, failure rate, correlation structure). Targeted coverage is 95%.

|            | Proposed |       |        |        |
| --- | --- | --- | --- | --- |
| (SS, FR, Cor) | MCP | LASSO | Oracle | Only.X |
| (450, 60%, Ind) | 0.955 | 0.922 | 0.950 | 0.790 |
| (300, 90%, Ind) | 0.938 | 0.935 | 0.968 | 0.935 |
| (300, 60%, Ind) | 0.943 | 0.838 | 0.940 | 0.863 |
| (450, 60%, AR) | 0.945 | 0.970 | 0.885 | 0.818 |
| (300, 90%, AR) | 0.922 | 0.955 | 0.945 | 0.897 |
| (300, 60%, AR) | 0.943 | 0.943 | 0.902 | 0.905 |
| (450, 60%, CS) | 0.965 | 0.968 | 0.940 | 0.938 |
| (300, 90%, CS) | 0.892 | 0.907 | 0.978 | 0.887 |
| (300, 60%, CS) | 0.910 | 0.917 | 0.897 | 0.938 |

having nonzero effects on survival. For nonzero regression coefficients, we took $\boldsymbol{\beta}_0 = (1,1)$ and $\boldsymbol{\theta}_{0A_0} = (1, 0.9, 0.8, 0.8, 0.9, 1)$. The 1,002 covariates were generated from a multivariate normal distribution with marginal means zero and marginal variances one. Three correlation scenarios were considered: (Ind) Under independence, all covariates are independent; (AR) Under the auto-regression correlation structure, the correlation coefficient between covariates $j$ and $k$ is $0.5^{|j-k|}$; (CS) Under the compound symmetry correlation structure, the correlation coefficient between covariates $j$ and $k$ is 0.1 if $j \neq k$. The event time was generated from the AFT model and followed a Weibull distribution. The censoring time was generated separately under two cases: (Case 1) the censoring time was exponential and independent of the covariates. Hence, also independent of the event time, and Assumption 1 is satisfied. (Case 2) The log censoring time was normal with mean satisfying the AFT model $C = X\boldsymbol{\beta}_0 + Z\boldsymbol{\theta}_1 + c_0$. The constant $c_0$ was used to adjust the censoring rate. The nonzero components of $\boldsymbol{\theta}_1$ took values $(1, 0.5, 0.5, 0.8, 0.9, 1)$. $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ shared three common nonzero components. Case 2 violates Assumption 1 and can serve the purpose of sensitivity analysis. For the (sample size, failure rate) dual, we considered (450, 60%), (300, 90%), and (300, 60%). Under each setting, 400 replicates were simulated.

The empirical coverage probabilities of the confidence regions are summarized in Tables 1 and 2. Taking all settings into consideration, overall, the proposed method, with both the MCP and LASSO penalties, generates satisfactory confidence regions with coverage probabilities close to the target. There are a small number of difficult scenarios, under which the proposed method does not behave very satisfactorily. However, under these scenarios, even the oracle

Table 2. Simulation: empirical coverage probability under Case 2. (SS, FR, Cor) = (sample size, failure rate, correlation structure). Targeted coverage is 95%.

| | Proposed | | | |
|---|---|---|---|---|
| (SS, FR, Cor) | MCP | LASSO | Oracle | Only.X |
| (450, 60%, Ind) | 0.988 | 0.980 | 0.993 | 1.000 |
| (300, 90%, Ind) | 0.910 | 0.955 | 0.953 | 0.990 |
| (300, 60%, Ind) | 0.980 | 0.945 | 0.985 | 0.960 |
| (450, 60%, AR) | 0.988 | 0.990 | 0.988 | 0.975 |
| (300, 90%, AR) | 0.922 | 0.953 | 0.960 | 0.998 |
| (300, 60%, AR) | 0.988 | 0.965 | 0.990 | 0.963 |
| (450, 60%, CS) | 0.995 | 0.985 | 0.998 | 0.993 |
| (300, 90%, CS) | 0.905 | 0.948 | 0.953 | 0.980 |
| (300, 60%, CS) | 0.990 | 0.968 | 0.990 | 0.990 |

method fails to deliver good results. The Only.X method, which focuses on the low-dimensional covariates, leads to unsatisfactory coverage. This is as expected. For each of the low-dimensional covariates, we also examine in Tables 3 and 4 the detailed marginal results. For the proposed and alternative methods, we computed the biases of estimates, standard errors, mean squared errors (MSE), and empirical coverage probabilities (CVR). In general, the proposed method generates competitive results, which are similar to those of the oracle method. Examining the MSEs suggests that with the proposed method, MCP may be preferred over LASSO. It is believed that this is caused by the better estimation performance of MCP over LASSO in general. The Only.X method is unsatisfactory, with biases and standard errors much larger than those of the other methods. The coverage probabilities in Tables 3 and 4 are not as satisfactory as those in Tables 1 and 2 (although the proposed method still has comparable performance as the oracle). The multivariate confidence regions in Tables 1 and 2 were accurately constructed using the Hotelling's $T^2$ statistics. In contrast, in Tables 3 and 4, a "brute force" approach was used to generate marginal confidence intervals based on the multivariate confidence regions. The marginal standard errors were taken directly from the diagonals of the estimated covariance matrices, and correlations among estimates were basically ignored. Similar problems have been noted in the literature and will not be discussed here. If the goal is to generate more accurate marginal confidence intervals, more sophisticated methods are needed. For example, one option is to construct the conditional confidence interval for each variable of interest given all other low-dimensional estimates.

Table 3. Simulation results for individual regression parameters under Case 1. All values are multiplied by 100 except for CVR.

| (SS, FR) | Method | $\beta_1 = 1$ Bias(se) | MSE | CVR | $\beta_2 = 1$ Bias(se) | MSE | CVR |
|---|---|---|---|---|---|---|---|
| | | Ind correlation | | | | | |
| | Proposed+MCP | 1.06 (3.82) | 22.34 | 0.968 | 0.77 (3.74) | 22.41 | 0.972 |
| (450, 60%) | Proposed+LASSO | −3.17 (4.04) | 32.23 | 0.958 | −3.79 (3.87) | 35.50 | 0.925 |
| | Oracle | 1.70 (3.96) | 25.84 | 0.950 | 1.56 (3.84) | 25.17 | 0.955 |
| | Only.X | −19.17(25.82) | 1,406.54 | 0.925 | −28.64(25.65) | 1,712.64 | 0.862 |
| | Proposed+MCP | 0.02 (2.37) | 8.68 | 0.948 | 0.78 (2.44) | 10.40 | 0.942 |
| (300, 90%) | Proposed+LASSO | −1.26 (2.53) | 10.29 | 0.908 | −1.14 (2.61) | 10.71 | 0.962 |
| | Oracle | 0.00 (2.33) | 7.97 | 0.975 | 0.47 (2.40) | 9.69 | 0.968 |
| | Only.X | −9.62(15.90) | 482.15 | 0.942 | −10.25(18.16) | 601.31 | 0.975 |
| | Proposed+MCP | 0.93 (4.04) | 29.25 | 0.938 | 1.31 (3.97) | 27.14 | 0.978 |
| (300, 60%) | Proposed+LASSO | −4.70 (5.07) | 62.38 | 0.888 | −5.87 (4.98) | 69.15 | 0.872 |
| | Oracle | 1.58 (4.16) | 31.36 | 0.938 | 2.06 (4.13) | 31.89 | 0.962 |
| | Only.X | −20.66(26.10) | 1,409.41 | 0.922 | −22.22(28.10) | 1,816.54 | 0.910 |
| | | AR correlation | | | | | |
| | Proposed+MCP | 1.67 (4.19) | 31.37 | 0.965 | 1.56 (4.48) | 31.99 | 0.988 |
| (450, 60%) | Proposed+LASSO | −1.41 (4.51) | 31.40 | 0.990 | −1.56 (4.49) | 36.10 | 0.972 |
| | Oracle | 2.15 (4.06) | 30.72 | 0.955 | 1.86 (4.07) | 29.88 | 0.978 |
| | Only.X | −19.39(28.88) | 1,562.74 | 0.945 | −18.01(28.28) | 1,555.93 | 0.925 |
| | Proposed+MCP | 0.81 (3.01) | 17.59 | 0.912 | 0.08 (3.29) | 18.04 | 0.945 |
| (300, 90%) | Proposed+LASSO | −0.05 (3.17) | 16.72 | 0.958 | −1.37 (3.29) | 18.48 | 0.945 |
| | Oracle | 0.80 (2.80) | 13.34 | 0.930 | 0.35 (2.77) | 11.79 | 0.962 |
| | Only.X | −4.95(20.70) | 660.11 | 0.955 | −9.14(20.20) | 679.58 | 0.930 |
| | Proposed+MCP | 1.56 (4.83) | 38.56 | 0.982 | 1.60 (4.89) | 42.25 | 0.955 |
| (300, 60%) | Proposed+LASSO | −3.89 (5.89) | 64.19 | 0.962 | −2.60 (5.84) | 57.70 | 0.952 |
| | Oracle | 2.13 (4.75) | 38.34 | 0.962 | 1.90 (4.66) | 41.56 | 0.932 |
| | Only.X | −22.27(29.00) | 1,734.95 | 0.932 | −14.38(28.72) | 1,395.78 | 0.975 |
| | | CS correlation | | | | | |
| | Proposed+MCP | 1.21 (3.61) | 21.63 | 0.982 | 1.30 (3.75) | 22.43 | 0.968 |
| (450, 60%) | Proposed+LASSO | −1.72 (3.92) | 26.67 | 0.982 | −2.11 (4.03) | 30.26 | 0.965 |
| | Oracle | 1.71 (3.66) | 22.76 | 0.962 | 1.72 (3.74) | 24.47 | 0.962 |
| | Only.X | −15.80(25.84) | 1,039.64 | 0.988 | −16.88(25.45) | 1,223.55 | 0.952 |
| | Proposed+MCP | −0.10 (2.54) | 11.00 | 0.950 | 0.56 (2.62) | 12.75 | 0.882 |
| (300, 90%) | Proposed+LASSO | −1.16 (2.67) | 12.51 | 0.920 | −1.13 (2.68) | 11.35 | 0.958 |
| | Oracle | −0.18 (2.46) | 9.41 | 0.962 | 0.19 (2.40) | 8.92 | 0.950 |
| | Only.X | −4.11(16.81) | 577.59 | 0.875 | −9.20(17.11) | 597.53 | 0.970 |
| | Proposed+MCP | 1.36 (3.90) | 25.96 | 0.970 | 2.16 (3.85) | 30.93 | 0.930 |
| (300, 60%) | Proposed+LASSO | −3.83 (4.95) | 57.59 | 0.898 | −2.45 (5.24) | 47.10 | 0.965 |
| | Oracle | 1.95 (3.95) | 27.65 | 0.955 | 2.69 (4.01) | 33.39 | 0.915 |
| | Only.X | −16.53(25.14) | 1,240.66 | 0.925 | −14.59(27.86) | 1,322.56 | 0.995 |

Table 4. Simulation results for individual regression parameters under Case 2. All values are multiplied by 100 except for CVR.

| (SS, FR) | Method | $\beta_1 = 1$ | | | $\beta_2 = 1$ | | |
|---|---|---|---|---|---|---|---|
| | | Bias(se) | MSE | CVR | Bias(se) | MSE | CVR |
| | | *Ind correlation* | | | | | |
| | Proposed+MCP | −0.19 (2.68) | 10.04 | 0.978 | −0.50 (2.69) | 10.04 | 0.990 |
| (450, 60%) | Proposed+LASSO | −0.55 (2.86) | 11.47 | 0.975 | −0.79 (2.86) | 12.01 | 0.978 |
| | Oracle | −0.18 (2.72) | 10.38 | 0.978 | −0.57 (2.72) | 10.06 | 0.990 |
| | Only.X | −2.61(17.48) | 392.39 | 0.995 | −1.08(16.81) | 391.15 | 0.995 |
| | Proposed+MCP | 0.37 (2.19) | 7.85 | 0.920 | 0.28 (2.29) | 8.30 | 0.915 |
| (300, 90%) | Proposed+LASSO | −0.06 (2.32) | 8.30 | 0.945 | 0.14 (2.36) | 8.49 | 0.970 |
| | Oracle | 0.31 (2.16) | 7.21 | 0.955 | 0.10 (2.17) | 6.91 | 0.972 |
| | Only.X | −1.22(13.94) | 283.02 | 0.990 | 0.22(13.59) | 289.44 | 0.930 |
| | Proposed+MCP | −0.47 (3.24) | 14.25 | 0.985 | −0.54 (3.17) | 14.63 | 0.978 |
| (300, 60%) | Proposed+LASSO | −0.89 (3.52) | 17.87 | 0.962 | −1.06 (3.43) | 18.89 | 0.938 |
| | Oracle | −0.41 (3.34) | 15.09 | 0.982 | −0.58 (3.28) | 15.45 | 0.985 |
| | Only.X | 0.56(21.03) | 644.01 | 0.938 | 0.15(20.73) | 618.00 | 0.968 |
| | | *AR correlation* | | | | | |
| | Proposed+MCP | −0.43 (3.14) | 14.32 | 0.985 | −0.07 (3.57) | 17.33 | 0.995 |
| (450, 60%) | Proposed+LASSO | −1.11 (3.51) | 17.87 | 0.970 | −0.16 (3.72) | 18.83 | 0.992 |
| | Oracle | −0.38 (3.15) | 13.65 | 0.985 | −0.42 (3.13) | 13.65 | 0.985 |
| | Only.X | −5.71(20.06) | 556.40 | 0.935 | 0.44(20.08) | 545.61 | 0.998 |
| | Proposed+MCP | −0.79 (2.72) | 15.30 | 0.922 | 0.44 (3.20) | 18.25 | 0.908 |
| (300, 90%) | Proposed+LASSO | −0.65 (2.90) | 13.69 | 0.962 | −0.29 (3.15) | 14.94 | 0.960 |
| | Oracle | −0.47 (2.53) | 10.32 | 0.960 | 0.29 (2.60) | 10.25 | 0.962 |
| | Only.X | 2.08(16.26) | 386.17 | 0.998 | −2.66(16.92) | 355.34 | 0.985 |
| | Proposed+MCP | −0.60 (3.84) | 21.09 | 0.985 | 0.30 (4.15) | 24.61 | 0.980 |
| (300, 60%) | Proposed+LASSO | −0.70 (4.29) | 26.53 | 0.968 | −0.27 (4.52) | 28.86 | 0.962 |
| | Oracle | −0.38 (3.82) | 20.62 | 0.992 | −0.19 (3.88) | 20.52 | 0.990 |
| | Only.X | 0.20(24.91) | 873.92 | 0.972 | 0.91(24.97) | 933.29 | 0.935 |
| | | *CS correlation* | | | | | |
| | Proposed+MCP | −0.22 (2.70) | 9.82 | 0.988 | −0.04 (2.65) | 9.83 | 0.990 |
| (450, 60%) | Proposed+LASSO | −0.24 (2.88) | 11.60 | 0.978 | −0.31 (2.85) | 10.99 | 0.982 |
| | Oracle | −0.14 (2.71) | 10.15 | 0.985 | −0.01 (2.69) | 9.89 | 0.980 |
| | Only.X | 1.49(17.34) | 398.15 | 0.982 | −1.30(17.35) | 384.94 | 0.985 |
| | Proposed+MCP | −0.11 (2.33) | 10.17 | 0.915 | 0.07 (2.35) | 9.78 | 0.912 |
| (300, 90%) | Proposed+LASSO | −0.37 (2.37) | 8.47 | 0.952 | −0.74 (2.45) | 9.08 | 0.938 |
| | Oracle | −0.18 (2.19) | 7.08 | 0.958 | −0.07 (2.24) | 7.50 | 0.940 |
| | Only.X | 1.64(13.74) | 322.17 | 0.965 | −5.84(14.52) | 315.59 | 0.995 |
| | Proposed+MCP | −0.14 (3.21) | 14.78 | 0.982 | 0.24 (3.22) | 14.39 | 0.982 |
| (300, 60%) | Proposed+LASSO | −1.17 (3.53) | 19.36 | 0.962 | −0.27 (3.49) | 17.66 | 0.958 |
| | Oracle | −0.17 (3.29) | 15.19 | 0.995 | 0.25 (3.28) | 14.82 | 0.990 |
| | Only.X | −4.79(21.06) | 650.7 | 0.990 | 2.37(20.46) | 674.77 | 0.988 |

## 5. Data Analysis

### 5.1. Analysis of the acute myeloid leukemia data

Data were retrieved from TCGA (The Cancer Genome Atlas; `https://tcga-data.nci.nih.gov/tcga/`), which is one of the most comprehensive genetic studies on cancer. For multiple cancers, prognosis data have been collected along with low-dimensional clinical measurements and high-dimensional genetic measurements. Acute myeloid leukemia (AML) is a cancer of the myeloid line of blood cells. It is the most common acute leukemia affecting adults, and its incidence increases with age. The dataset we analyzed contains 194 records. The event rate was 66.8%. The response variable was overall survival. There are three clinical variables of interest ($\mathbf{X}$): FAB category (which is a classification system that runs from M0 through to M7, with M0 being the best and M7 being the worst. Here we assume that the severity of disease is linear in FAB.), age at initial pathologic diagnosis, and white blood cell count. Data on the expressions of 19,798 genes are available. To improve stability, we conducted a supervised screening prior to analysis. Specifically, we first conducted marginal analysis and computed the univariate correlation coefficient between the response and each gene expression. Then genes with p-values (for the correlation coefficients) less than 0.1 were selected. These selected genes were then ordered by their interquartile ranges, and the top 50% were selected. Overall, 1,001 gene expressions were further analyzed.

We applied the three methods as in the previous section. For the low-dimensional covariates, the estimates (multiplied by $10^2$) and estimated covariance matrices (multiplied by $10^4$) were $\hat{\boldsymbol{\beta}}_{Only.X} = (86.73, 6.70, 0.41)^\top, \hat{\boldsymbol{\beta}}_{LASSO} = (8.10, 2.83, 4.28)^\top, \hat{\boldsymbol{\beta}}_{MCP} = (31.53, 1.08, 7.56)^\top,$

$$\hat{\boldsymbol{\Sigma}}_{Only.X} = \begin{pmatrix} 219.97 & -11.21 & -1.89 \\ -11.21 & 0.92 & -0.07 \\ -1.89 & -0.07 & 0.32 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_{LASSO} = \begin{pmatrix} 265.13 & -8.80 & -13.88 \\ -8.80 & 1.08 & -0.58 \\ -13.88 & -0.58 & 2.68 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_{MCP} = \begin{pmatrix} 437.22 & -27.44 & -5.66 \\ -27.44 & 5.54 & -5.28 \\ -5.66 & -5.28 & 9.75 \end{pmatrix}.$$

The level $1 - \alpha$ confidence region can be constructed as $D_* = \left\{ \boldsymbol{\beta} \in \mathbb{R}^p : \left(\hat{\boldsymbol{\beta}}_* - \right.\right.$

Table 5. Analysis of the AML data: 95% marginal confidence intervals for the low-dimensional covariates (results multiplied by 100) and lists of identified genes.

|  | FAB category | Age | WBC |
|---|---|---|---|
| Proposed+MCP | [−9.452,  72.513] | [−3.535, 5.694] | [1.438, 13.676] |
| Proposed+LASSO | [−23.809,  40.018] | [0.795, 4.863] | [1.070,  7.484] |
| Only.X | [57.665, 115.802] | [4.828, 8.580] | [−0.695,  1.513] |
| | Identified genes | | |
| Proposed+MCP | TPPP, OXCT1, TSPAN2, C10orf81, RHCG, SOCS1, TMEM132E, FCGBP, LOC613126, ZGLP1, GDF2 | | |
| Proposed+LASSO | LOC100131508, TPPP, OXCT1, TLR3, TSPAN2, C10orf81, RHCG, LIPC, KIT, SOCS1, TMEM132E, PCOLCE, GDF2, FCGBP, CHRNA6, OPCML, LOC613126, NYX, ZGLP1, LMTK3, CD109, TECTB, PRPH, HOPX, SLC7A9, LOC100130331 | | |

$\boldsymbol{\beta}\big)^{\top}\hat{\boldsymbol{\Sigma}}_*^{-1}\big(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}\big) < 193 \times 3/191 F_{1-\alpha,3,191}\big\}$, where $* \in \{Only.X, LASSO, MCP\}$. The marginal confidence intervals are provided in Table 5. Different methods generate different results. For example, FAB is significant and white blood cell count is not under the low-dimensional model. The observations are reversed under the proposed method. With the proposed method, LASSO and MCP identify different sets of genes. In addition, the significance level of age differs under the two penalties.

## 5.2. Analysis of the glioblastoma data

Glioblastoma multiforme (GBM) is tumor that arises from astrocytes. It is the most common and most aggressive malignant primary brain tumor in human. The dataset analyzed was obtained from TCGA and contains 298 records. The response variable is overall survival, and the event rate is 68.7%. There are four low-dimensional clinical variables of interest: gender (binary, male is coded as 1), race (binary, white is coded as 1), Karnofsky score (a quality of life measure that runs from 0 to 100, with 0 being the worst), and age at initial pathologic diagnosis. Measurements on 17,800 gene expressions are available. We conducted similar marginal screening as in the previous data analysis and selected 1,188 gene expressions for downstream analysis.

We applied the proposed and alternative methods. For the low-dimensional covariates, the estimates (multiplied by $10^2$) and estimated covariance matrices (multiplied by $10^4$) were

$$\hat{\boldsymbol{\beta}}_{Only.X} = (-1.47, 23.11, 6.23, 1.91)^{\top},$$
$$\hat{\boldsymbol{\beta}}_{LASSO} = (-18.78, 36.25, 6.31, 1.77)^{\top},$$

Table 6. Analysis of the glioblastoma data: 95% marginal confidence intervals for the low-dimensional covariates. Results are multiplied by 100.

|  | Gender | Race | Karnofsky | Age |
|---|---|---|---|---|
| Proposed+MCP | $[-64.801, 26.392]$ | $[-64.741, 127.434]$ | $[4.970, 7.699]$ | $[0.406, 3.143]$ |
| Proposed+LASSO | $[-63.798, 26.231]$ | $[-60.826, 133.325]$ | $[4.989, 7.621]$ | $[0.471, 3.071]$ |
| Only.X | $[-30.365, 27.429]$ | $[-23.771, \; 69.997]$ | $[5.596, 6.867]$ | $[1.147, 2.674]$ |

$$\hat{\boldsymbol{\beta}}_{MCP} = (-19.20, 31.35, 6.33, 1.77)^{\top},$$

$$\hat{\boldsymbol{\Sigma}}_{Only.X} = \begin{pmatrix} 217.37 & -13.85 & -0.92 & -1.03 \\ -13.85 & 572.21 & -3.90 & -3.75 \\ -0.92 & -3.90 & 0.11 & -0.06 \\ -1.03 & -3.75 & -0.06 & 0.15 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_{LASSO} = \begin{pmatrix} 527.49 & 257.16 & -6.75 & -2.09 \\ 257.16 & 2453.16 & -22.31 & -13.47 \\ -6.75 & -22.31 & 0.45 & -0.14 \\ -2.09 & -13.47 & -0.14 & 0.44 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_{MCP} = \begin{pmatrix} 541.22 & 270.10 & -7.14 & -1.93 \\ 270.10 & 2403.47 & -22.13 & -12.99 \\ -7.14 & -22.13 & 0.48 & -0.18 \\ -1.93 & -12.99 & -0.18 & 0.49 \end{pmatrix}.$$

Confidence regions can be constructed as described above. In Table 6, we provide the marginal 95% confidence intervals. With the proposed method, the LASSO and MCP penalties generate reasonably similar results, different from Only.X. In addition, genes FSIP1, NRBP2 and CHST10 are identified by the proposed method with LASSO penalty, while gene CHST10 is identified with the MCP penalty.

## 5.3. Remarks

With the proposed method, both LASSO and MCP can be theoretically valid. In practical data analysis, it is observed that the two penalties may lead to different results. Simulation suggests that the MCP results may be preferred. We note that in simulation, discrepancies are also observed, however, much smaller than in practical data analysis. It is believed that this is caused by the noisier nature of practical data. We have conducted small-scale simulations and found that, if other settings are kept unchanged and sample size is increased, then the differences between LASSO and MCP results can shrink (results omitted). With

the complexity of the proposed analysis and unknown data/model settings in practice, it is difficult to determine what is the minimum sample size needed in order to generate similar LASSO and MCP results. Variable selection is not of main interest in our analysis. It is expected that, with their differences in general, LASSO and MCP may generate different variable selection, as observed in data analysis.

## 6. Discussion

In this study, we have considered censored survival data under the AFT model where the covariate effects can be partitioned into two components. Data that have a small set of covariates of special interest are commonly encountered. This study thus has practical implications. We have developed an approach that can conduct inference for the low-dimensional covariates. Establishment of its theoretical properties not only provides a strong ground for the proposed method, but also sheds lights into high-dimensional inference in general. With the presence of censoring and the AFT model, this study complements and advances existing literature. Simulations show satisfactory finite-sample performance. The analysis of two datasets demonstrates practical applicability of the method.

This study can be potentially extended in multiple ways. For accommodating censoring in the estimation of AFT model, there exist other techniques, for example the Buckley-James and rank-based. These techniques are also of interest but not chosen in this study because of their high computational cost. It may also be of interest to extend the analysis of AFT model to other survival models. Statistical inference under high-dimensional settings is a fast-developing field. It would be of interest to compare different methods in the future.

## Supplementary Materials

This file includes proofs of the theoretical results presented in the main text.

## Acknowledgment

# References

Bae, J. and Kim, S. (2003). The uniform central limit theorem for the Kaplan-Meier integral process. *Bulletin of the Australian Mathematical Society* **67**, 467–480.

Belloni, A., Chernozhukov, V. and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* **81**, 608–650.

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics* **41**, 802–837.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19**, 1212–1242.

Cai, T., Huang, J. and Tian, L. (2009). Regularized estimation for the accelerated failure time model. *Biometrics* **65**, 394–404.

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when $p$ is much larger than $n$. *The Annals of Statistics* **35**, 2313–2351.

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

Fang, E. X., Ning, Y. and Liu, H. (2016). Testing and confidence intervals for high dimensional proportional hazards model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Hu, J. and Chai, H. (2013). Adjusted regularized estimation in the accelerated failure time model with high dimensional covariates. *Journal of Multivariate Analysis* **122**, 96–114.

Huang, J. and Ma, S. (2010). Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Analysis* **16**, 176–195.

Huang, J., Ma, S. and Xie, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62**, 813–820.

Huang, J., Ma, S. and Xie, H. (2007). Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica* **17**, 1533–1548.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research* **15**, 2869–2909.

Johnson, B. A. (2008). Variable selection in semiparametric linear regression with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 351–370.

Johnson, B. A. (2009). Rank-based estimation in the l1-regularized partly linear model for censored outcomes with application to integrated analyses of clinical predictors and gene expression data. *Biostatistics (Oxford, England)* **10**, 659–666.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* **186**, 325–344.

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016). Exact post-selection inference, with application to the Lasso. *The Annals of Statistics* **44**, 907–927.

Ma, S. and Du, P. (2012). Variable selection in partly linear regression model with diverging dimensions for right censored data. *Statistica Sinica* **22**, 1003–1020.

Morris, V. K., Lucas, F. A. S., Overman, M. J., Eng, C., Morelli, M. P., Jiang, Z.-Q., Luthra, R., Meric-Bernstam, F., Maru, D., Scheet, P., Kopetz, S. and Vilar, E. (2014). Clinico-pathologic characteristics and gene expression analyses of non-KRAS 12/13, RAS-mutated metastatic colorectal cancer. *Annals of Oncology* **25**, 2008–2014.

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics* **45**, 158–195.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89–103.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics* **23**, 461–471.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodological)* **58**, 267–288.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics* **42**, 1166–1202.

van de Geer, S. and Lederer, J. (2013). The bernstein–orlicz norm and deviation inequalities. *Probability Theory and Related Fields* **157**, 225–250.

van de Geer, S. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics* **3**, 1360–1392.

van der Vaart, A. W. and Wellner, J. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer: New York.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge university press: Cambridge.

Voorman, A., Shojaie, A. and Witten, D. (2014). Inference in high dimensions with the penalized score test. *ArXiv E-Prints* ArXiv: 1401.2678.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

Yang, Z., Wang, Z., Liu, H., Eldar, Y. C. and Zhang, T. (2016). Sparse nonlinear regression: Parameter estimation and asymptotic inference. *International Conference on Machine Learning*. 1–32.

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**, 894–942.

Zhang, C. H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 217–242.

Department of Biostatistics, 60 College ST, New Haven, CT 06520 USA.

E-mail: haochai2@gmail.com

School of Economics, Xiamen University, Siming District, Xiamen, Fujian Province, China, 361005.

E-mail: qzzhang.wise@gmail.com

241 Schaeffer Hall, University of Iowa, Iowa City, IA 52242-1419, USA.

E-mail: jian-huang@uiowa.edu

Department of Biostatistics, 60 College ST, New Haven CT, 06520 USA.

E-mail: shuangge.ma@yale.edu