*The Annals of Statistics*
2018, Vol. 46, No. 1, 308–343
https://doi.org/10.1214/17-AOS1551


# HIGH DIMENSIONAL CENSORED QUANTILE REGRESSION

BY QI ZHENG[1], LIMIN PENG[2] AND XUMING HE[3]

*University of Louisville, Emory University and University of Michigan*

Censored quantile regression (CQR) has emerged as a useful regression tool for survival analysis. Some commonly used CQR methods can be characterized by stochastic integral-based estimating equations in a sequential manner across quantile levels. In this paper, we analyze CQR in a high dimensional setting where the regression functions over a continuum of quantile levels are of interest. We propose a two-step penalization procedure, which accommodates stochastic integral based estimating equations and address the challenges due to the recursive nature of the procedure. We establish the uniform convergence rates for the proposed estimators, and investigate the properties on weak convergence and variable selection. We conduct numerical studies to confirm our theoretical findings and illustrate the practical utility of our proposals.

**1. Introduction.** High dimensional data arise in a wide variety of scientific studies such as genomics, neuroimaging, social network and finance. In such data, the number of candidate covariates $p$ often greatly exceeds the number of observations $n$, but the number of truly relevant variables $s$ is relatively small. This poses unprecedented challenges and opportunities in statistical analysis.

Penalization methods have been intensively studied to tackle these challenges. A number of penalty functions such as Lasso [Tibshirani (1996)], SCAD [Fan and Li (2001)], Adaptive Lasso (ALasso [Zou (2006)]), and MCP [Zhang (2010)] have been developed and investigated with various regression models. However, the related development in dealing with censored survival (i.e., time-to-event) responses has been relatively sparse. Most of the existing approaches focus on the Cox proportional hazard model and the accelerated failure time model (AFT) [Huang, Ma and Xie (2006), Johnson (2009), Huang and Ma (2010), Bradic, Fan and Jiang (2011), among others]. Despite the success of the Cox model and the AFT model in survival analysis, their limitations from assuming constant covariate effects (e.g., proportional hazards, location shift effects) have been recognized. For example,

Received May 2016; revised January 2017.
[1]Supported by NIH Grant R01HL113548.
[2]Supported by NIH Grants R01HL113548 and R01MH079448.
[3]Supported by NSF Awards, DMS-13-07566 and DMS-16-07840, and National Natural Science Foundation of China Grant 11129101.
*MSC2010 subject classifications.* Primary 62J07; secondary 62H12.
*Key words and phrases.* High dimensional survival data, varying covariate effects, censored quantile regression.

variable selection based on these models may pose a considerable risk of missing variables that have complex, nonconstant survival impact but are scientifically important [Peng, Xu and Kutner (2014)].

Quantile regression [Koenker and Bassett (1978)] has emerged as a useful alternative regression strategy for survival analysis. By allowing covariates to have varying impact across different segments of the response distribution, quantile regression significantly extends the traditional AFT regression. Quantile regression also provides a direct approach to predicting the quantiles of a survival response (which have better identifiability than the mean in the presence of censoring). These nice modeling features make quantile regression a desirable platform for investigating covariate sparsity and building parsimonious prediction models for a survival response subjects to censoring.

Censored quantile regression (CQR) methods that deal with randomly censored survival data have been well studied in the finite $p$ case [Peng and Huang (2008), Powell (1986), Ying, Jung and Wei (1995), Portnoy (2003), Wang and Wang (2009), among others]. Two popular CQR approaches are Portnoy's (2003) method and its variants [Neocleous, Vanden Branden and Portnoy (2006), Portnoy and Lin (2010)], which were grounded on the principle of self-consistency [Efron (1967)], and Peng and Huang's (2008) method, which was built upon the martingale structure of randomly censored data. These methods require the relatively weak random censoring assumption and enjoy efficient and stable implementation by standard statistical software, for example, R function $crq(\cdot)$ [Koenker (2016)], and SAS procedure PROC QUANTREG. With $p$ fixed, several authors also investigated the variable selection problem for CQR. For example, Shows, Lu and Zhang (2010) and Peng, Xu and Kutner (2014) proposed penalized estimating equations where censoring was handled by inverse probability weighting. Wang, Zhou and Li (2013) developed a robust variable selection method based on Wang and Wang's (2009) work, adopting a global dimension reduction formulation to facilitate the local weight estimation. Volgushev, Wagener and Dette (2014) studied an extension of Peng and Huang's (2008) method with properly designed penalty terms. The existing work on fixed $p$ problems provides useful insights for developing high dimensional CQR methods.

In this work, we study the high dimensional CQR methods based on Portnoy (2003) (and its variants) and Peng and Huang (2008). We focus on these methods, because they require weaker assumptions on censoring but use a recursive scheme in estimation, which poses technical challenges. As delineated in Peng (2012), the CQR approach proposed by either Portnoy (2003) or Peng and Huang (2008) is essentially attached to a stochastic integral based estimating equation. As a result, the estimation is not separately done at each individual quantile index but performed sequentially across quantile indices with the estimate corresponding to a larger quantile index depending on estimates obtained for lower quantile indices. With fixed $p$, traditional empirical process arguments, coupled with induction, were used by Peng and Huang (2008) and Volgushev, Wagener and Dette

(2014) to control the errors cumulated from sequentially estimating the stochastic process of the quantile index (i.e., cumulative estimation errors). However, such a technique does not apply directly when $p > n$. To overcome this difficulty, we apply the generic chaining idea from Talagrand (2005) and van de Geer (2008) and incorporate it into a sophisticated induction framework that fits both sequential estimation and penalization. As shown in our theoretical studies, there exists a complicated entanglement between the penalization and the control of cumulative estimation errors. This fact necessitates a dynamic penalization scheme, which is new in the literature. Our work constitutes the first effort to tackle high dimensional stochastic integral based estimating equations, and the techniques employed here can be useful in other settings.

In this paper, we adopt the perspective of globally concerned quantile regression [Zheng, Peng and He (2015)], which concerns conditional quantiles over a continuum of quantile indices. Compared to the classical practice of quantile regression, which is generally confined to model a single or multiple quantiles at fixed quantile levels [e.g., Fan, Fan and Barut (2014), Wang, Wu and Li (2012), Zheng, Gallagher and Kulasekera (2013)], the globally concerned quantile regression enjoys some advantages: (1) it utilizes all useful information across quantiles to improve the robustness of variable selection; (2) it grasps global sparsity in a more concise way; (3) it also encompasses locally concerned quantile regression [see Zheng, Peng and He (2015)].

Our proposed procedure on high dimensional CQR consists of two steps. In the first step, we incorporate a Lasso type $L_1$ penalty into the stochastic integral based estimating equation for CQR and obtain a uniformly consistent estimator. In the second step, we use an ALasso type penalty to reduce the bias induced by the $L_1$ penalty and the resulting estimator achieves improved estimation efficiency and model selection consistency.

The first step of our procedure is not a routine implementation of $L_1$ penalization. The penalization at a quantile level $\tau$ not only has to take into account the local sparsity, but also needs to adapt stochastic process estimation errors resulted from penalization at other $\tau$'s. Note that adopting the $L_1$ penalties in this step helps us avoid the nonconvex optimization issues, and at the same time facilitates the adjustment for cumulative estimation errors by maintaining the same forms of penalty across quantiles. With well-modulated penalties, we can establish a sharp upper bound for local estimation error at each $\tau$. Consequently, we are able to derive a useful formula for the upper bound of the cumulative estimation error, which renders a uniform convergence rate $\sqrt{s\log(p\vee n)/n}$ for the resulting estimator of the coefficient function. This result lays the foundation for the subsequent ALasso step.

With ALasso penalties weighted by the estimator obtained from the first step, we can reduce the estimation bias to the order $\sqrt{s\log n/n}$, the oracle rate of uniform convergence. In this step, the cumulative estimation errors become negligible compared to the amount of penalization that ensures sparsity. This enables us to adopt a uniform selector of the tuning parameter to control the overall model complexity over the set of quantile levels of interest. In the development of the tuning parameter selector, to accommodate censoring, we propose to measure model fitness by calibrating the violation from the assumed martingale process by using deviance residuals. Along this line, we develop a general information criterion (GIC) type tuning parameter selector, which allows us to achieve model selection consistency and also performs well in finite-sample simulation studies. In addition, we also show that our proposed ALasso penalized estimator converges weakly to a Gaussian process as a function of the quantile index $\tau$.

The rest of the article is organized as follows. In Section 2, we briefly review the finite dimensional CQR methods and the basic framework of globally concerned CQR. In Section 3, we introduce our two-step penalized CQR method with Lasso type and Adaptive Lasso type penalties. The theoretical properties of our proposed estimator are investigated in Section 4. We conduct simulation studies to evaluate the finite sample performance of the proposed estimators in Section 5. In Section 5, an application to a real data example is also presented. We defer all technical proofs and some discussions to the Appendix.

## 2. Preliminaries.

2.1. *Censored quantile regression.* Let $T$ denote a survival response of interest, which is subject to right censoring by $C$. Under the standard random censoring assumption, $C$ is independent of $T$ given $\tilde{\mathbf{Z}}$, a $(p-1)\times 1$ covariate vector $(p>1)$. Let $X=\min\{T,C\}$, $\Delta = I(T\le C)$, and $\mathbf{Z}=(1,\tilde{\mathbf{Z}}^T)^T$. The observed data consist of $n$ i.i.d replicates of $(X,\Delta,\mathbf{Z})$, denoted by $\{(X_i,\Delta_i,\mathbf{Z}_i), i=1,\dots,n\}$.

Define the $\tau$th conditional quantile of $T$ given $\mathbf{Z}$ as $Q_T(\tau|\mathbf{Z})=\inf\{t:\Pr(T\le t|\mathbf{Z})\ge\tau\}$. A linear quantile regression model for $\log T$ may take the form

$$Q_T(\tau|\mathbf{Z})=\exp\{\mathbf{Z}^T\boldsymbol{\beta}_0(\tau)\},\qquad \tau\in(0,\tau_U], \tag{1}$$

where $\boldsymbol{\beta}_0(\tau)$ is a $p$-dimensional vector of unknown coefficients and $0<\tau_U<1$. The nonintercept coefficients in $\boldsymbol{\beta}_0(\tau)$ represent the effects of the covariates on the $\tau$th quantile of $\log T$ given $\mathbf{Z}$. It is worth noting that model (1) is confined to a $\tau$-interval away from 1, $(0,\tau_U]$ with $0<\tau_U<1$, due to the identifiability issue at upper quantiles due to censoring. Peng and Huang (2008) discussed the theoretical constraints on $\tau_U$ as well as the practical selection of $\tau_U$ with real datasets.

Peng and Huang (2008) proposed a stochastic integral based estimating equation for $\boldsymbol{\beta}_0(\tau)$ in model (1). More recently, Peng (2012) showed that the self-consistent CQR approaches [Neocleous, Vanden Branden and Portnoy (2006), Portnoy (2003), Portnoy and Lin (2010), for example] can also be formulated as stochastic integral based estimating equations. For the sake of simplicity, we shall focus on Peng and Huang's (2008) method for CQR in the sequel.

Let $N_i(t)=1\{\log X_i\le t,\Delta_i=1\}$, $\Lambda_T(t|\mathbf{Z})=-\log(1-\Pr(\log T\le t|\mathbf{Z}))$, and $H(u)=-\log(1-u)$. By Fleming and Harrington (1991), $M_i(t):=N_i(t)-$

$\Lambda_T(t \wedge \log X_i|\mathbf{Z}_i)$ is a martingale process, and hence $E[M_i(t)|\mathbf{Z}_i] = 0$. This, combined with the fact that $\Lambda_T(\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau) \wedge \log X_i|\mathbf{Z}_i) = \int_0^\tau 1\{\log X_i \geq \mathbf{Z}_i^T\boldsymbol{\beta}_0(u)\}\,dH(u)$ under model (1), implies

$$E\left[\mathbf{Z}_i\left(N_i(\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau)) - \int_0^\tau 1\{\log X_i \geq \mathbf{Z}_i^T\boldsymbol{\beta}_0(u)\}\,dH(u)\right)\right] = 0 \quad \text{for } 0 < \tau \leq \tau_U. \tag{2}$$

Motivated by this fact, Peng and Huang (2008) proposed to estimate $\boldsymbol{\beta}_0(\cdot)$ by the following estimating equation:

$$n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i\left[N_i(\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau)) - \int_0^\tau 1\{\log X_i \geq \mathbf{Z}_i^T\boldsymbol{\beta}_0(u)\}\,dH(u)\right] = 0 \quad \tau \in (0, \tau_U]. \tag{3}$$

The way that equation (3) involves stochastic integrals entails a sequential estimation procedure for $\boldsymbol{\beta}_0(\tau)$, $\tau \in (0, \tau_U]$. More specifically, let $\Gamma_{m_n}$ be a grid of $\tau$-values, $0 = \tau_0 < \tau_1 < \cdots < \tau_{m_n} = \tau_U$. Peng and Huang's (2008) estimator of $\boldsymbol{\beta}_0(\tau)$, denoted by $\check{\boldsymbol{\beta}}(\tau)$, is defined as a right-continuous piecewise-constant function that only jumps at the grid points, satisfying $\exp(\mathbf{Z}_i^T\check{\boldsymbol{\beta}}(\tau_0)) = 0$ for all $1 \leq i \leq n$. At grid points greater than $\tau_0$, $\check{\boldsymbol{\beta}}(\tau_k)$, $k = 1, \ldots, m_n$, are sequentially obtained by solving the following estimating equation for $\mathbf{h}$:

$$n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i\left(N_i(\mathbf{Z}_i^T\mathbf{h}) - \sum_{r=0}^{k-1}\int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i \geq \mathbf{Z}_i^T\check{\boldsymbol{\beta}}(\tau_r)\}\,dH(u)\right) = 0. \tag{4}$$

Equation (4) is a monotone estimating equation and can be efficiently solved via $L_1$-minimization. Peng and Huang (2008) showed that $\check{\boldsymbol{\beta}}(\tau)$ is uniformly consistent and converges weakly to a mean zero Gaussian process for $\tau \in [\varpi, \tau_U]$, where $0 < \varpi < \tau_U$.

2.2. *Globally concerned CQR framework.* Much of the published work on high dimensional quantile regression is generally confined to model a single or multiple quantiles [Wang, Wu and Li (2012), Zheng, Gallagher and Kulasekera (2013), Fan, Fan and Barut (2014), for example]. Referred to as locally concerned quantile regression, this strategy involves choices of the fixed quantile levels, which might be natural in some applications, but highly subjective in others.

In this paper, we take the perspective of globally concerned quantile regression [Zheng, Peng and He (2015)] to tackle the high dimensional modeling of CQR. That is, we specify an interval of quantile indices $[\tau_L, \tau_U]$ according to scientific questions of interest. Our goal under high dimensional CQR (HDCQR) is to identify the set of relevant variables, defined as

$$S^* := \bigcup_{\tau \in [\tau_L, \tau_U]} \operatorname{supp}(\boldsymbol{\beta}_0(\tau)) = \left\{j : \sup_{\tau \in [\tau_L, \tau_U]} |\beta_0^{(j)}(\tau)| \neq 0\right\},$$

and then to estimate $\boldsymbol{\beta}_0^{(j)}(\tau)$ for $j \in S^*$ and $\tau \in [\tau_L, \tau_U]$. Here and hereafter, $u^{(j)}$ denotes the $j$th element of the vector $\mathbf{u}$ and supp($\mathbf{u}$) denotes $\{j : u^{(j)} \neq 0\}$. Studying the HDCQR under this set-up will render information on which variables and how they influence the targeted segment of the distribution of the survival response.

## 3. Proposed methods.

3.1. *HDCQR estimator with Lasso type penalties* (*L-HDCQR*). It is clear that equation (4) for the low dimensional CQR is not directly solvable when $p \gg n$. We use Lasso type penalties to address this problem.

We first introduce the following assumption to circumvent the singularity problem with censored regression quantile at $\tau = 0$.

ASSUMPTION 3.1. There exists a quantile index $\upsilon$ and some constant $c$ such that $n^{-1} \sum_{i=1}^{n} 1\{\log C_i \leq \mathbf{Z}_i^T \boldsymbol{\beta}_0(\upsilon)\}(1 - \Delta_i) \leq cn^{-1/2}$ holds for sufficiently large $n$.

This assumption requires that the number of censored observations below the $\upsilon$th quantile does not exceed $cn^{1/2}$. A similar but stronger condition is imposed in Portnoy (2003), which requires no censoring below the $\upsilon$th quantile. Note that $\mathbf{Z}_i^T \boldsymbol{\beta}_0(0) = -\infty$ under model (1). Therefore, Assumption 3.1 can be met when the lower bound of $C$'s support is greater than 0, a situation that seems reasonable for most censoring mechanisms in survival analysis. We cannot really verify Assumption 3.1 for any given data set, because the assumption is asymptotic in nature, but it does suggest how we need to choose $\upsilon$ in real data analysis. Based on our numerical experiences, we recommend choosing $\upsilon$ as a value such that only a small proportion of the $X_i's$ below the fitted $\upsilon$th quantiles are censored. One may further confirm the selection of $\upsilon$ by conducting a sensitivity analysis that repeats the estimation with different choices of $\upsilon$. Our simulations in Section 5.1 suggest that as long as $\upsilon$ is chosen reasonably small (e.g., $\leq 0.1$), the proposed method has a robust performance even when this technical assumption is violated with a positive probability.

Under Assumption 3.1, simple calculations yield that $E[\int_0^\upsilon 1\{\log X_i \geq \mathbf{Z}_i^T \boldsymbol{\beta}_0(u)\} \, dH(u)] = \upsilon$. Consequently, we obtain a slightly modified version of equation (2):

$$E\left[\mathbf{Z}_i\left(N_i(\mathbf{Z}_i^T \boldsymbol{\beta}_0(\tau)) - \int_\upsilon^\tau 1\{\log X_i \geq \mathbf{Z}_i^T \boldsymbol{\beta}_0(u)\} \, dH(u) - \upsilon\right)\right] = 0 \quad \text{for } \upsilon \leq \tau \leq \tau_U. \tag{5}$$

The modification in (5) to the stochastic integral implies that one can start the sequential estimation procedure from the $\upsilon$th quantile. Therefore, throughout the rest of the paper, we set $\tau_0 = \upsilon$ in the $\tau$-grid $\Gamma_{m_n}$. Moreover, we require $\Gamma_{m_n}$ to be a equally spaced grid for simplicity. A more general choice of the grid is discussed in the Section B of the supplementary material [Zheng, Peng and He (2018)]; see Remark 2.

Without assuming any prior information about the impact of covariates on the $\tau$th quantile of $T$, we adopt Lasso type penalties, and the proposed L-HDCQR estimator, $\tilde{\boldsymbol{\beta}}(\cdot)$, is defined as a right-continuous piecewise-constant function that only jumps on the grid points of $\Gamma_{m_n}$. At each $\tau_k \in \Gamma_{m_n}$, $\tilde{\boldsymbol{\beta}}(\tau_k)$ is obtained as the minimizer of the following $L_1$-type convex objective functions sequentially:

$$\tilde{Q}_k(\mathbf{h}) = \tilde{L}_k(\mathbf{h}) + \tilde{\lambda}_{k,n}\|\mathbf{h}\|_1, \qquad 0 \le k \le m_n, \tag{6}$$

where $\tilde{\lambda}_{k,n}$ is the tuning parameter, and

$$\begin{aligned}\tilde{L}_k(\mathbf{h}) = {} & \sum_{i=1}^n \Delta_i \left|\log X_i - \mathbf{h}^T \mathbf{Z}_i\right| + \mathbf{h}^T \sum_{i=1}^n \Delta_i \mathbf{Z}_i \\ & - 2\mathbf{h}^T \sum_{i=1}^n \mathbf{Z}_i \left(\sum_{r=0}^{k-1} \int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i \ge \mathbf{Z}_i^T \tilde{\boldsymbol{\beta}}(\tau_r)\}\, dH(u) + \tau_0\right).\end{aligned} \tag{7}$$

In particular, $\tilde{L}_0(\mathbf{h}) = \sum_{i=1}^n \Delta_i |\log X_i - \mathbf{h}^T \mathbf{Z}_i| + \mathbf{h}^T \sum_{i=1}^n \Delta_i \mathbf{Z}_i - 2\mathbf{h}^T \sum_{i=1}^n \mathbf{Z}_i \tau_0$. For any $\tau_k < \tau < \tau_{k+1}$, $0 \le k \le m_n - 1$, we set $\tilde{\boldsymbol{\beta}}(\tau) = \tilde{\boldsymbol{\beta}}(\tau_k)$.

It is worth mentioning that locating the minimizers of (6) is equivalent to finding the generalized solutions to the following equations:

$$\sum_{i=1}^n \mathbf{Z}_i \big(N_i(\mathbf{Z}_i^T \mathbf{h}) - \tau_0\big) + \frac{1}{2}\tilde{\lambda}_{0,n}\, \mathrm{sign}(\mathbf{h}) \approx 0, \tag{8}$$

$$\begin{aligned}& \sum_{i=1}^n \mathbf{Z}_i \left(N_i(\mathbf{Z}_i^T \mathbf{h}) - \sum_{r=0}^{k-1} \int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i \ge \mathbf{Z}_i^T \tilde{\boldsymbol{\beta}}(\tau_r)\}\, dH(u)\right) \\ & \quad + \frac{1}{2}\tilde{\lambda}_{k,n}\, \mathrm{sign}(\mathbf{h}) \approx 0, \qquad k \ge 1.\end{aligned} \tag{9}$$

These estimating equations are the first-order equivalents of (6), and can be viewed as a penalized version of the stochastic integral based estimating equation (5).

The minimization problem for (6) can be easily solved by using **R** package *quantreg*. Specifically, let $\Delta_{n+1} = \Delta_{n+2} = 1$ and $\log X_{n+1} = \log X_{n+2} = R$, where $R$ is a large constant. For example, we set $R = 10^4$ in our numerical studies. Let $\mathbf{Z}_i(k) = \mathbf{Z}_i, i = 1, \dots, n$, $\mathbf{Z}_{n+1}(k) = 2\sum_{i=1}^n \mathbf{Z}_i(\sum_{r=0}^{k-1} \int_{\tau_r}^{\tau_{r+1}}, 1\{\log X_i \ge \mathbf{Z}_i^T \tilde{\boldsymbol{\beta}}(\tau_r)\}\, dH(u) + \tau_0)$, and $\mathbf{Z}_{n+2}(k) = -\sum_i^n \Delta_i \mathbf{Z}_i$. Then the $\tilde{Q}_k(\mathbf{h})$ in (6) can be re-expressed as $\sum_{i=1}^{n+2} |\Delta_i \log X_i - \mathbf{h}^T \Delta_i \mathbf{Z}_i(k)| + \tilde{\lambda}_{k,n}\|\mathbf{h}\|_1$. This suggests that the minimization of $\tilde{Q}_k(\mathbf{h})$ can be solved by penalized uncensored quantile regression with responses $\{\Delta_i \log X_i\}_{i=1}^{n+2}$, covariates $\{\Delta_i \mathbf{Z}_i(k)\}_{i=1}^{n+2}$ and tuning parameter $\tilde{\lambda}_{k,n}$, which can implemented by the $rq(\cdot)$ function in **R**.

3.2. *HDCQR estimator with ALasso type penalties* (*AL-HDCQR*). The theoretical results in Section 4.2 suggest that the proposed L-HDCQR estimator is uniformly consistent, but its convergence rate is sub-optimal. It also does not produce consistent variable selection. To improve on the HDCQR we consider the HDCQR estimator with weighted Lasso penalty, which has been demonstrated as an effective tool to ameliorate the estimation bias and improve variable selection accuracy.

The basic strategy is to modify the L-HDCQR procedure by assigning weighted penalties to covariates according to their impact on the response distribution, which can be captured by the proposed L-HDCQR estimator $\tilde{\boldsymbol{\beta}}(\tau)$. Following the work by Zheng, Peng and He (2015), the weighted penalties, $\omega^{(j)}(\tau)$ ($\tau \in [\tau_L, \tau_U]$, $j = 2, \dots, p$), under the globally concerned CQR, can be specified in the following ways:

(a) (Pointwise) $\omega^{(j)}(\tau) = \chi(|\tilde{\beta}^{(j)}(\tau)|)$;

(b) (Average) $\omega^{(j)}(\tau) = \chi(\int_{\tau_0}^{\tau_L} |\tilde{\beta}^{(j)}(u)|\,du)$ if $\tau_0 \le \tau < \tau_L$, and $\omega^{(j)}(\tau) = \chi(\int_{\tau_L}^{\tau_U} |\tilde{\beta}^{(j)}(u)|du)$ if $\tau_L \le \tau \le \tau_U$;

(c) (Uniform) $\omega^{(j)}(\tau) = \chi(\sup_{\tau_0 \le \tau < \tau_L} |\tilde{\beta}^{(j)}(\tau)|)$ if $\tau_0 \le \tau < \tau_L$, and $\omega^{(j)}(\tau) = \chi(\sup_{\tau_L \le \tau \le \tau_U} |\tilde{\beta}^{(j)}(\tau)|)$ if $\tau_L \le \tau \le \tau_U$;

where $\chi(\cdot)$ is a nonincreasing function. The pointwise weight in (a) corresponds to a traditional way of choosing adaptive weight, which captures the "local" importance of a covariate at a single $\tau$. The average weight in (b) and the uniform weight in (c) are tailored to the globally concerned quantile regression framework and are designed to penalize a covariate according to its cumulative effect or maximum effect over a specified $\tau$-interval. Note that in (b) and (c), the average weight and the uniform weight are assigned different values for $\tau \in [\tau_0, \tau_L)$ and $\tau \in [\tau_L, \tau_U]$.

Denote the proposed AL-HDCQR estimator by $\hat{\boldsymbol{\beta}}(\tau)$. We can obtain $\hat{\boldsymbol{\beta}}(\tau_k)$'s by sequentially minimizing

$$\hat{Q}_k(\mathbf{h}) = \hat{L}_k(\mathbf{h}) + \lambda_{n*}\|\boldsymbol{\omega}_k \circ \mathbf{h}\|_1 \qquad \text{for all } \tau_0 \le \tau_k < \tau_L, \tag{10}$$

$$\hat{Q}_k(\mathbf{h}) = \hat{L}_k(\mathbf{h}) + \lambda_n^*\|\boldsymbol{\omega}_k \circ \mathbf{h}\|_1 \qquad \text{for all } \tau_L \le \tau_k \le \tau_U, \tag{11}$$

where

$$\begin{aligned}\hat{L}_k(\mathbf{h}) = {} & \sum_{i=1}^n \Delta_i \big|\log X_i - \mathbf{h}^T \mathbf{Z}_i\big| + \mathbf{h}^T \sum_{i=1}^n \Delta_i \mathbf{Z}_i \\ & - 2\mathbf{h}^T \sum_{i=1}^n \mathbf{Z}_i \left( \sum_{r=0}^{k-1} \int_{\tau_r}^{\tau_{r+1}} 1\big\{\log X_i \ge \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}(\tau_r)\big\}\, dH(\tau) + \tau_0 \right),\end{aligned}$$

$\lambda_{n*}$ and $\lambda_n^*$ are tuning parameters for $[\tau_0, \tau_L)$ and $[\tau_L, \tau_U]$, respectively, $\boldsymbol{\omega}_k$ is the vector function $(0, \omega^{(2)}(\tau), \dots, \omega^{(p)}(\tau))^T$ evaluated at $\tau = \tau_k$, and $\circ$ denotes the Hadamard product. Following the discussion on the minimization of (6) in

Section 3.1, one can obtain the minimizers of (10) and (11) similarly by using the $rq(\cdot)$ function in **R**. A detailed description of our computational algorithm for obtaining $\hat{\boldsymbol{\beta}}(\tau)$ is provided in Section F of the supplemental article [Zheng, Peng and He (2018)].

As shown above, the proposed AL-HDCQR estimator adopts different tuning parameters and possibly different adaptive weights for the penalization over $\tau \in [\tau_0, \tau_L)$ and $\tau \in [\tau_L, \tau_U]$. Doing so allows us to tailor the shrinkage of $\{\hat{\boldsymbol{\beta}}(\tau) : \tau \in [\tau_L, \tau_U]\}$ to the sparsity over $\tau \in [\tau_L, \tau_U]$, the quantile levels of interest in $S*$. The penalization over $\tau \in [\tau_0, \tau_L)$ is aimed at providing consistent estimates for $\{\boldsymbol{\beta}_0(\tau) : \tau \in [\tau_0, \tau_L)\}$ that have estimation errors bounded by $O(\sqrt{s \log n/n})$. This may be achieved by penalization strategies different from that shown in (10). For the sake of simplicity in the presentation, we adopt the same type of penalization in each set of $\tau \in [\tau_0, \tau_L)$ and $\tau \in [\tau_L, \tau_U]$.

It is worth emphasizing that the tuning parameter $\lambda_n^*$ is constant over $[\tau_L, \tau_U]$. We are allowed to do so, because the cumulative estimation errors in the AL-HDCQR are significantly reduced by the introduction of adaptive weights, and thus they would not affect the sparsity, if $\lambda_n^*$ and $\boldsymbol{\omega}$ are chosen appropriately. Using a common $\lambda_n^*$ for $\tau \in [\tau_L, \tau_U]$ allows us to tune for the sparsity across the targeted $\tau$-region rather than local sparsity at each $\tau$. As noted in Zheng, Peng and He (2015), this plays an important role in achieving consistent variable selection by the proposed AL-HDCQR. Theoretical constraints on $\lambda_n^*$ are provided in Propositions 4.1 and 4.2, and Theorem 4.3. We will discuss the selections of tuning parameters in Section 4.4.

There are also many different choices for the function $\chi(\cdot)$. For example, we can choose $\chi_1(u) = 1/u$, for $u > 0$ [Zou (2006)], $\chi_2(u) = 1\{u \le \lambda_n/n\} + 1\{u > \lambda_n/n\}(a\lambda_n/n - u)_+/((a-1)\lambda_n/n)$ for a given constant $a > 2$ and tuning parameter $\lambda_n$ [Fan and Li (2001)], or $\chi_3(u) = (1 - u/(\gamma\lambda_n/n))_+$, for a given constant $\gamma > 0$ [Zhang (2010)]. In our numerical studies, we use the SCAD function $\chi_2(u)$ with $a = 2.4$.

## 4. Theoretical studies.

4.1. *Notation and regularity conditions.* Given a random sample $Z_1, \dots, Z_n$, let $\mathbb{G}_n(f) = \mathbb{G}_n(f(Z_i)) := n^{-1/2}\sum_{i=1}^n (f(Z_i) - E[f(Z_i)])$ and $\mathbb{E}_n f = \mathbb{E}_n f(Z_i) := \sum_{i=1}^n f(Z_i)/n$. We use $\|\cdot\|_r$ to denote the $l_r$-norm. In particular, we denote the $l_2$-norm by $\|\cdot\|$. Given a vector $\boldsymbol{\delta} \in \mathbb{R}^p$ and a set of indices $T \subset \{1, \dots, p\}$, we denote by $T^c$ and $|T|$ the complementary set and the cardinality of $T$ respectively, and denote by $\boldsymbol{\delta}_T$ the vector in which $\boldsymbol{\delta}_T^{(j)} = \boldsymbol{\delta}^{(j)}$ if $j \in T$, and $\boldsymbol{\delta}_T^{(j)} = 0$ if $j \notin T$. Define $\|\boldsymbol{\delta}\|_{r,T} = \|\boldsymbol{\delta}_T\|_r$. The loss function of quantile regression at $\tau$th quantile is denoted by $\rho_\tau(u) := u(\tau - 1\{u < 0\})$.

*Regularity conditions*: We assume the following regularity conditions:

(C1) (*Condition on covariates*): $\|\mathbf{Z}\|_\infty \le C_0$, for some constant $C_0$. Without loss of generality, we assume $C_0 = 1$.

(C2) (*Condition on the conditional density*): Let $F_T(t|\mathbf{Z}) = \Pr(\log T \le t|\mathbf{Z})$, $\Lambda_T(t|\mathbf{Z}) = -\log(1 - F_T(t|\mathbf{Z}))$, $F(t|\mathbf{Z}) = \Pr(\log X \le t|\mathbf{Z})$, and $G(t|\mathbf{Z}) = \Pr(\log X \le t, \Delta = 1|\mathbf{Z})$. Also, define $f(t|\mathbf{Z}) = dF(t|\mathbf{Z})/dt$, and $g(t|\mathbf{Z}) = dG(t|\mathbf{Z})/dt$.

(a) There exist constants $\underline{f}$, $\bar{f}$, $\underline{g}$ and $\bar{g}$ such that

$$\underline{f} \le \inf_{\mathbf{z},\tau\in[\tau_0,\tau_U]} f(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z}) \le \sup_{\mathbf{z},\tau\in[\tau_0,\tau_U]} f(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z}) \le \bar{f},$$

$$\underline{g} \le \inf_{\mathbf{z},\tau\in[\tau_0,\tau_U]} g(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z}) \le \sup_{\mathbf{z},\tau\in[\tau_0,\tau_U]} g(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z}) \le \bar{g}.$$

(b) There exist constants $\kappa$ and $A$, such that $\forall |t| \le \kappa$,

$$\sup_{\mathbf{z},\tau\in[\tau_0,\tau_U]} \left| f(\mathbf{z}^T\boldsymbol{\beta}_0(\tau) + t|\mathbf{z}) - f(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z})\right| \le A|t|,$$

$$\sup_{\mathbf{z},\tau\in[\tau_0,\tau_U]} \left| g(\mathbf{z}^T\boldsymbol{\beta}_0(\tau) + t|\mathbf{z}) - g(\mathbf{z}^T\boldsymbol{\beta}_0(\tau)|\mathbf{z})\right| \le A|t|.$$

(C3) (*Sparsity and dimensionality*): Let $S_\tau := \{j : |\boldsymbol{\beta}_0^{(j)}(\tau)| > 0, 1 \le j \le p\}$, $S_* := \bigcup_{\tau\in[\tau_0,\tau_L)} S_\tau = \{\sup_{\tau_0\le\tau<\tau_L} |\beta_0^{(j)}(\tau)| \ne 0\}$, the set of relevant covariates over $[\tau_0, \tau_L)$, and $S := \bigcup_{\tau\in[\tau_0,\tau_U]} S_\tau = S_* \cup S^*$, where $S^*$ stands for the set of relevant covariates over $[\tau_L, \tau_U]$ as defined in Section 2.2. We assume $\log p = o(n^{1/2})$ and $s := |S|$ does not change with $n$.

(C4) [*Smoothness of* $\boldsymbol{\beta}_0(\tau)$]: let $\tilde{\mu}(\tau) = E[1\{\log X > \mathbf{Z}^T\boldsymbol{\beta}_0(\tau)\}]$. There exist a positive constant $L$, such that $|\beta_0^{(j)}(\tau_1) - \beta_0^{(j)}(\tau_2)| \le L|\tau_1 - \tau_2|$ and $|\tilde{\mu}(\tau_1) - \tilde{\mu}(\tau_2)| < L|\tau_1 - \tau_2|$, for all $\tau_1, \tau_2 \in [\tau_0, \tau_U]$ and $1 \le j \le p$.

(C5) (*Restricted eigenvalue condition and restricted nonlinear impact*): Let $A_\tau$ denote the restricted set $\{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_{1,S_\tau^c} \le ((c_0 + 1)/(c_0 - 1))\|\boldsymbol{\delta}\|_{1,S_\tau}$, $\|\boldsymbol{\delta}\|_{0,S_\tau^c} \le n\}$ and $A_S$ denote $\{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_{1,S^c} \le ((c_0 + 1)/(c_0 - 1))\|\boldsymbol{\delta}\|_{1,S}$, $\|\boldsymbol{\delta}\|_{0,S^c} \le n\}$, for some constant $c_0 > 1$. We can see that $A_\tau \subset A_S$, for all $\tau \in [\tau_0, \tau_U]$.

(a) (Restricted eigenvalue condition) $0 < \lambda_{\min} \le \inf_{\boldsymbol{\delta}\in A_S,\boldsymbol{\delta}\ne\mathbf{0}} \boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}/\|\boldsymbol{\delta}\|^2$.

(b) (Restricted nonlinear impact) $q := \inf_{\boldsymbol{\delta}\in A_S,\boldsymbol{\delta}\ne\mathbf{0}} E[(\mathbf{Z}_i^T\boldsymbol{\delta})^2]^{3/2}/E[|\mathbf{Z}_i^T\boldsymbol{\delta}|^3] > 0$.

(C6) (Grid size of $\Gamma_{m_n}$): Let $\epsilon_n := \tau_k - \tau_{k-1}, \tau_k \in \Gamma_{m_n}$. The grid size satisfies $cn^{-1} \le \epsilon_n \le c^{-1}\sqrt{\log(p \wedge n)/n}$ for some constant $c$.

Condition (C1) assumes the boundedness of the covariates, which is reasonable in most applications. Following the discussions in Koenker (2005), Zheng, Peng and He (2015) pointed out that a global linear quantile regression model is most sensible when the covariates are confined to a compact set. In an unbounded covariate space, the linear quantile functions have to be parallel, which means i.i.d. errors; otherwise the quantiles functions would cross. The same argument applies to censored quantile regression. Under a location shift linear model, we may allow the support of the covariates to be unbounded. This is confirmed by the simulation results reported in the supplemental article [Zheng, Peng and He (2018)]. Condition (C2) imposes mild assumptions on the conditional density $f(t|\mathbf{Z}) \equiv d\Pr(\log T \wedge \log C \le t|\mathbf{Z})/dt$, which, by its definition, accounts for the effect of censoring on the observed data. Condition (C2) implies the positiveness of $f(t|\mathbf{Z})$ everywhere between $\mathbf{Z}^T\boldsymbol{\beta}_0(\tau_L)$ and $\mathbf{Z}^T\boldsymbol{\beta}_0(\tau_U)$, which plays an important role to ensure the identifiability of $\boldsymbol{\beta}_0(\tau)$ for $\tau < \tau_U$. The selection of $\tau_U$ in real data analysis can follow the same principle suggested by Peng and Huang (2008) for finite-dimensional CQR. By Condition (C3), the number of candidate covariates may increase at an exponential rate as a function of $n$, but the true model size $s$ is fixed over $n$. In fact, Condition (C3) can be relaxed to allow $s$ to increase with $n$. More specifically, we show in Theorem 4.1 that

$$\sup_{\tau_0 \le \tau \le \tau_U} \|\tilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \le \big(C_1 \exp(C_2 \tau_U s) + L \cdot c^{-1}\big)\sqrt{s \log(p \vee n)/n},$$

which implies that our proposed estimator is still consistent if $s$ increases at the rate $o(\log n)$. However, $s$ is not allowed to grow at the rate of $o(n^{1/3})$ as in He and Shao (2000) due to the unique challenge in controlling the cumulative estimation error in our estimation procedure. Condition (C4) requires that $\boldsymbol{\beta}_0(\tau)$ is sufficiently smooth. Condition (C5) is a type of assumptions commonly seen in the high dimensional data analysis literature [Belloni and Chernozhukov (2011), Bickel, Ritov and Tsybakov (2009), Fan, Fan and Barut (2014), for example]. The eigenvalue condition (C5a) is crucial to the model identifiability. The restricted nonlinear impact (RNI) coefficient $q$ in condition (C5b) controls the quality of minoration of the quantile loss function by a quadratic function, which is needed to establish the consistency of the proposed estimators. Condition (C6) imposes constraints on the fineness of the grid $\Gamma_{m_n}$. The condition $cn^{-1} \le \epsilon_n$ is required merely to simplify our proof. We can relax it with more tedious arguments.

4.2. *Theoretical properties of L-HDCQR.* The sequential nature of our L-HDCQR estimation procedure suggests investigating the properties of $\tilde{\boldsymbol{\beta}}(\tau)$ by induction.

Let $\{\nu_{k,n}(b), k = 0, \dots, m_n\}$ be a positive sequence satisfying $\nu_{0,n}(b) = \nu_{0,n} = \sqrt{s \log(p \vee n)/n}$ and $\nu_{k+1,n}(b) = \nu_{k,n}(1 + bs\epsilon_n)$ for some constant $b$, where $\epsilon_n := \tau_k - \tau_{k-1}$. Given two positive constants $a$ and $b$, we define $\tilde{\Omega}_k(a, b)$, $k = 0, \dots, m_n$ as the event that for all $0 \le r \le k$,

$$\inf_{\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}\boldsymbol{\delta}\| = \sqrt{\lambda_{\min}} a \nu_{r,n}(b), \boldsymbol{\delta} \in A_{\tau_r}} \tilde{Q}_r\big(\boldsymbol{\beta}_0(\tau_r) + \boldsymbol{\delta}\big) - \tilde{Q}_r\big(\boldsymbol{\beta}_0(\tau_r)\big) > 0 \quad \text{and}$$

$$\tilde{\boldsymbol{\beta}}(\tau_r) - \boldsymbol{\beta}_0(\tau_r) \in A_{\tau_r},$$

where $A_{\tau_r} = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}\|_{1,S_{\tau_r}^c} \le ((c_0+1)/(c_0-1))\|\boldsymbol{\delta}\|_{1,S_{\tau_r}}, \|\boldsymbol{\delta}\|_{0,S_{\tau_r}^c} \le n\}$ is defined in Condition (C5). Note that $\tilde{\Omega}_0(a,b)$ only depends on $a$ since $\nu_{0,n}(b)$ is free of $b$ by its definition. It can be seen that if the event $\tilde{\Omega}_k(a,b)$ holds, then by the convexity of $\tilde{Q}_r(\mathbf{h})$, $\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}(\tilde{\boldsymbol{\beta}}(\tau_r) - \boldsymbol{\beta}_0(\tau_r))\| \le \sqrt{\lambda_{\min}} a\nu_{r,n}(b)$, and consequently $\|\tilde{\boldsymbol{\beta}}(\tau_r) - \boldsymbol{\beta}_0(\tau_r)\| \le a\nu_{r,n}(b)$, for all $0 \le r \le k$, according to Condition (C5a). Thus, $\tilde{\Omega}_k(a,b)$ can be roughly viewed as the event that $\tilde{\boldsymbol{\beta}}(\tau_r)$ is located in the local ball with center $\boldsymbol{\beta}_0(\tau_r)$ and radius $a\nu_{r,n}(b)$ for all $0 \le r \le k$. If we can show that there exist some constants $a$ and $b$, such that $\Pr(\tilde{\Omega}_{m_n}(a,b)) \to 1$, and $\nu_{m_n,n}(b) \to 0$, then the uniform estimation consistency of $\tilde{\boldsymbol{\beta}}(\tau)$ for $\tau \in [\tau_0, \tau_U]$ can be established given the smoothness of $\boldsymbol{\beta}_0(\tau_0)$ imposed by Condition (C3).

We apply induction arguments to examine the probability that $\tilde{\Omega}_k(a,b)$ holds. Specifically, in Proposition A.1 (see Appendix A), we show that there exists a universal large constant $C_1$, such that $P(\tilde{\Omega}_0(C_1,0)) > 1 - 16\exp(-4\log(p\wedge n)) - 2\exp(-3\log(p\vee n))$. As discussed above, this establishes the estimation consistency of $\tilde{\boldsymbol{\beta}}(\tau_0)$ with the convergence rate $\sqrt{s\log(p\vee n)/n}$, and hence provides the baseline result for the induction. Next, we establish the key result for the induction in Proposition A.2 (see Appendix A). That is, there exists a universal large constant $C_2$, such that given the event $\tilde{\Omega}_{k-1}(C_1, C_2)$ holds, $k = 1,\dots,m_n$, if the tuning parameter $\tilde{\lambda}_{k,n}$ and $\epsilon_n$ are well chosen, then the event $\tilde{\Omega}_k(C_1,C_2)$ holds with probability at least $1 - 4(5k+7)\exp(-3\log(p\vee n))$. It is worthy mentioning here that in Proposition A.2 we show that the cumulative estimation errors from $\tilde{\boldsymbol{\beta}}(\tau_r)'s, r < k$ involved in (7) may achieve the sparsity threshold level $\sqrt{n\log(p\vee n)}$ and deteriorate the estimate at $\tau_k$, thus we need to adjust the tuning parameter to produce sparse consistent estimates. More details are presented in Appendix A.

By Propositions A.1 and A.2, we can infer the estimation consistency of the L-HDCQR estimator over $[\tau_0, \tau_U]$. First, according to the definition of the sequence that $\nu_{k+1,n}(C_2) \le \nu_{k,n}(C_2)(1 + C_2 s\epsilon_n), 0 \le k \le m_n - 1$, we have $\nu_{m_n,n}(C_2) \le \nu_{0,n}(1 + C_2 s\epsilon_n)^{\tau_U/\epsilon_n} \to \exp(C_2 s\tau_U)\sqrt{s\log(p\vee n)/n}$, which converges to 0 with the rate $\sqrt{s\log(p\vee n)/n}$, under Condition (C3). Next, by our choices of $\epsilon_n$ from Condition (C6) for grid $\Gamma_{m_n}$, we obtain that $\Pr(\tilde{\Omega}_{m_n}(C_1,C_2)) \ge 1 - \sum_{k=0}^{m_n} 4(5k + 7)\exp(-3\log(p\vee n)) \ge 1 - C_3(p\vee n)^{-1}$ for some constant $C_3$. Combining the above two results yields the estimation consistency of $\tilde{\boldsymbol{\beta}}(\tau_k)$ at $k = 1,\dots,m_n$, and consequently by Condition (C4), we can establish the uniform consistency of $\tilde{\boldsymbol{\beta}}(\tau)$ over $[\tau_0, \tau_U]$ as stated in the following theorem.

THEOREM 4.1. *Suppose the conditions in Propositions A.1 and A.2 are satisfied. There exist some constants $C_1$, $C_2$ and $C_3$, such that*

$$\sup_{\tau_0\le\tau\le\tau_U} \|\tilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \le C_1\exp(C_2 s\tau_U)\sqrt{s\log(p\vee n)/n} + L\sqrt{s}\epsilon_n$$
$$\le \big(C_1\exp(C_2 s\tau_U) + L\cdot c^{-1}\big)\sqrt{s\log(p\vee n)/n},$$

*with probability at least* $1 - C_3(p \vee n)^{-1}$, *where* $L$ *and* $c$ *are defined in Conditions* (C4) *and* (C6), *respectively.*

The model identified by $\tilde{\boldsymbol{\beta}}(\tau)$ over $[\tau_L, \tau_U]$ is denoted by $\tilde{S}^* := \{j : \sup_{\tau\in[\tau_L,\tau_U]} |\tilde{\beta}^{(j)}(\tau)| \neq 0\}$. Based on the uniform consistency of $\tilde{\boldsymbol{\beta}}(\tau)$, we can see that if the maximum signal of each relevant covariate $j \in S^*$ is strong enough in $[\tau_L, \tau_U]$ such that $\sup_{\tau\in[\tau_L,\tau_U]} |\tilde{\beta}^{(j)}(\tau)| \geq \sup_{\tau\in[\tau_L,\tau_U]} |\beta_0^{(j)}(\tau)| - (C_1 \exp(C_2\tau_U s) + L \cdot c^{-1})\sqrt{s\log(p \vee n)/n} > 0$, then $S^*$, the true model over $[\tau_L, \tau_U]$ is included in $\tilde{S}^*$, as stated by the following corollary.

COROLLARY 4.1. *Under Conditions* (C1)–(C6) *and Assumption* 3.1, *if* $\sup_{\tau\in[\tau_L,\tau_U]} |\beta_0^{(j)}(\tau)| > C_1 \exp(C_2 s\tau_U)\sqrt{s\log(p \vee n)/n} + L\sqrt{s}\epsilon_n$, *for all* $j \in S^*$, *then* $S^* \subseteq \tilde{S}^*$ *with probability at least* $1 - C_3(p \vee n)^{-1}$. *Moreover, provided any hard-thresholding* $\gamma$, *such that* $(C_1 \exp(C_2\tau_U s) + L \cdot c^{-1})\sqrt{s\log(p \vee n)/n} < \gamma < \inf_{j\in S^*} \sup_{\tau_L\leq\tau\leq\tau_U} |\beta_0^{(j)}(\tau)| - (C_1 \exp(C_2\tau_U s) + L \cdot c^{-1})\sqrt{s\log(p \vee n)/n}$,

$$\Big\{j : \sup_{\tau_L\leq\tau\leq\tau_U} |\tilde{\beta}^{(j)}(\tau)| > \gamma\Big\} = S^*,$$

*with probability at least* $1 - C_3(p \vee n)^{-1}$.

The second result of Corollary 4.1 indicates that if the maximum signals of all relevant covariates are beyond $2(C_1 \exp(C_2\tau_U s) + L \cdot c^{-1})\sqrt{s\log(p \vee n)/n}$, the additional hard thresholding can screen out all irrelevant covariates that are mis-selected by $\tilde{\boldsymbol{\beta}}(\tau)$ over $[\tau_L, \tau_U]$. However, the hard thresholding here explicitly depends on the unknown constants $C_1$ and $C_2$, and hence has little practical utility. Moreover, as shown in the literature [Zou (2006), Zheng, Gallagher and Kulasekera (2013), Fan, Fan and Barut (2014), for example] and in our numerical analysis, the Lasso-type estimators usually suffer from large bias, especially in high dimensional data.

4.3. *Theoretical properties of AL-HDCQR*. Let $\{\vartheta_{k,n}(b), k = 0, \ldots, m_n\}$ be a positive sequence satisfying $\vartheta_{0,n}(b) = \vartheta_{0,n} = \sqrt{s\log n/n}$ and $\vartheta_{k+1,n}(b) = \vartheta_{k,n}(1 + bs\epsilon_n)$ for some constant $b$. Given some constants $a$ and $b$, we define $\hat{\Omega}_k(a, b)$, $k = 0, \ldots, m_n$ as the event that for all $0 \leq r \leq k$,

$$\inf_{\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}\boldsymbol{\delta}\| = \sqrt{\lambda_{\min}}a\vartheta_{r,n}(b), \operatorname{supp}(\boldsymbol{\delta})\subseteq S_*} \hat{Q}_r\big(\boldsymbol{\beta}_0(\tau_r) + \boldsymbol{\delta}\big) - \hat{Q}_r\big(\boldsymbol{\beta}_0(\tau_r)\big) > 0,$$

$$\big\{j : \hat{\boldsymbol{\beta}}^{(j)}(\tau_r) \neq 0\big\} \subseteq S_*$$

when $\tau_0 \leq \tau_r < \tau_L$, and

$$\inf_{\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}\boldsymbol{\delta}\| = \sqrt{\lambda_{\min}}a\vartheta_{r,n}(b), \operatorname{supp}(\boldsymbol{\delta})\subseteq S^*} \hat{Q}_r\big(\boldsymbol{\beta}_0(\tau_r) + \boldsymbol{\delta}\big) - \hat{Q}_r\big(\boldsymbol{\beta}_0(\tau_r)\big) > 0,$$

$$\big\{j : \hat{\boldsymbol{\beta}}^{(j)}(\tau_r) \neq 0\big\} \subseteq S^*$$

when $\tau_L \le \tau_r \le \tau_U$. Similar to $\tilde{\Omega}_k(a,b)$, $\hat{\Omega}_k(a,b)$ can be approximately viewed as the event that $\hat{\boldsymbol{\beta}}(\tau_r)$ is not only located in the local ball with center $\boldsymbol{\beta}_0(\tau_r)$ and radius $a\vartheta_{r,n}(b)$, but also does not include any irrelevant variables, for all $0 \le r \le k$. We would like to show that $\Pr(\hat{\Omega}_{m_n}(a,b)) \to 1$, and $\vartheta_{m_n,n}(b) \to 0$ for some constants $a$ and $b$, from which the estimation consistency of the AL-HDCQR estimator follows.

PROPOSITION 4.1. *Under Conditions* (C1)–(C6) *and Assumption* 3.1, *if* $\lambda_{n*}\|\boldsymbol{\omega}_0\|_{2,S_*} \le c_2\sqrt{sn\log n}$ *for some constant* $c_2$, *and* $\inf_{j\notin S_*}\lambda_{n*}\omega_0^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$, *then there exists a sufficiently large constant* $C_4$, *such that the event* $\hat{\Omega}_0(C_4, 0)$ *holds with probability at least* $1 - 38\exp(-3\log(p\vee n))$.

PROPOSITION 4.2. *Under Conditions* (C1)–(C6), *if* $\lambda_{n*}\|\boldsymbol{\omega}_k\|_{2,S_*} \le c_2\sqrt{sn\log n}$, $\lambda_n^*\|\boldsymbol{\omega}_k\|_{2,S^*} \le c_2\sqrt{sn\log n}$, *and* $\inf_{j\notin S_*}\lambda_{n*}\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$, $\inf_{j\notin S^*}\lambda_n^*\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$, *there exists a universal constant* $C_5$, *such that given the event* $\hat{\Omega}_{k-1}(C_4, C_5)$, *the event* $\hat{\Omega}_k(C_4, C_5)$ *holds with probability at least* $1 - 2(19k+27)\exp(-3\log(p\vee n))$.

The results in Propositions 4.1 and 4.2 are analogous to those in Propositions A.1 and A.2 in Appendix A, respectively. From Propositions 4.1 and 4.2, it can be seen that if the tuning parameters and adaptive weights are appropriately imposed such that the irrelevant covariates receive strong penalties over the level $\sqrt{sn\log(p\vee n)}$ and the penalties for relevant covariates are no more than $c_2\sqrt{sn\log n}$ at all $\tau_r < \tau_k$, then the estimation bias can be reduced to the level $\sqrt{s\log n/n}$. As a result, the errors from the part $\sum_{r=0}^{k-1}\int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i \ge \mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau_r)\}\,dH(u)$ in $\hat{L}_k(\mathbf{h})$ at $\tau_k$ is at most of the level $\sqrt{sn\log n}$, which is negligible compared to the sparsity threshold level for the irrelevant covariates, as $\inf_{j\notin S_*}\lambda_{n*}\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$ and $\inf_{j\notin S^*}\lambda_n^*\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$. Therefore, the sparsity would not be altered and it is unnecessary to adjust the tuning parameter and adaptive weights to control the cumulative estimation errors.

Propositions 4.1 and 4.2 imply that $\vartheta_{m_n,n}(C_5) \le \vartheta_{0,n}(1 + C_5 s\epsilon_n)^{\tau_U/\epsilon_n} \to C_4\exp(C_5 s\tau_U)\sqrt{s\log n/n}$, which converges to 0 with the rate $\sqrt{s\log n/n}$, and also $\Pr(\hat{\Omega}_{m_n}(C_4, C_5)) \ge 1 - \sum_{k=0}^{m_n} 2(19k+27)\exp(-3\log(p\vee n)) \ge 1 - C_6(p\vee n)^{-1}$ for some constant $C_6$. Then by Condition (C4), we obtain the uniform consistency of $\hat{\boldsymbol{\beta}}(\tau)$ across $\tau \in [\tau_0, \tau_U]$, as stated in the following theorem.

THEOREM 4.2. *Suppose the conditions in Propositions* 4.1 *and* 4.2 *are satisfied. Then there exist some constants* $C_4$, $C_5$, *and* $C_6$, *such that*

$$\sup_{\tau_0\le\tau\le\tau_U}\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \le \big(C_4\exp(C_5 s\tau_U) + L\cdot c^{-1}\big)\sqrt{s\log n/n},$$

*with probability at least* $1 - C_6(p\vee n)^{-1}$, *where* $L$ *is defined in Condition* (C4).

Noting that the uniform rate of $\hat{\boldsymbol{\beta}}(\tau)$ is improved to $\sqrt{s\log n/n}$, By the definition of $\hat{\Omega}_k(C_4, C_5)$, $\Pr(\hat{\Omega}_{m_n}(C_4, C_5)) \to 1$ also implies that the model identified by $\hat{\boldsymbol{\beta}}(\tau)$ over $[\tau_L, \tau_U]$, $\hat{S}^* := \bigcup_{\tau\in[\tau_L,\tau_U]}\{j : \hat{\beta}^{(j)}(\tau) \neq 0\}$, is a submodel of $S^*$ with probability tending to 1. Furthermore, if $\inf_{j\in S^*}\sup_{\tau\in[\tau_L,\tau_U]}|\beta^{(j)}(\tau)| > (C_4\exp(C_5 s\tau_U) + L\cdot c^{-1})\sqrt{s\log n/n}$, then $\hat{S}^* = S^*$ can be inferred from the estimation consistency of $\hat{\boldsymbol{\beta}}(\tau)$. The model selection consistency of the AL-HDCQR estimator is stated in the next corollary.

COROLLARY 4.2. *Suppose the conditions in Theorem 4.2 hold, if* $\inf_{j\in S^*}\sup_{\tau\in[\tau_L,\tau_U]}|\beta^{(j)}(\tau)| > (C_4\exp(C_5 s\tau_U) + L\cdot c^{-1})\sqrt{s\log n/n}$, *then* $\hat{S}^* = S^*$ *with probability at least* $1 - C_6(p\vee n)^{-1}$.

If we adopt the uniform weighted penalty $\omega_k^{(j)} = \chi_2(\sup_{\tau_0\le\tau_r<\tau_L}|\tilde{\beta}^{(j)}(\tau_r)|)$ for all $\tau_0 \le \tau_k < \tau_L$, and $\omega_k^{(j)} = \chi_2(\sup_{\tau_L\le\tau_r\le\tau_U}|\tilde{\beta}^{(j)}(\tau_r)|)$ for all $\tau_L \le \tau_k \le \tau_U$, $j = 2,\dots,p$, and $\chi_2(u) = 1\{u \le \lambda_n/n\} + 1\{u > \lambda_n/n\}(a\lambda_n/n - u)_+/((a-1)\lambda_n/n)$, then the conditions $\max_{1\le k\le m_n}\{\lambda_{n*}\|\boldsymbol{\omega}_k\|_{2,S_*}, \lambda_n^*\|\boldsymbol{\omega}_k\|_{2,S^*}\} \le c_2\sqrt{sn\log n}$ and $\min_{1\le k\le m_n}\{\inf_{j\notin S_*}\lambda_{n*}\omega_k^{(j)}, \inf_{j\notin S^*}\lambda_n^*\omega_k^{(j)}\}/\sqrt{sn\log(p\vee n)} \to \infty$ can be satisfied by choosing

$$\lambda_{n*}, \lambda_n^* = O\bigl(\sqrt{sn\log(p\vee n)\log\log n}\bigr),$$

$$\inf_{j\in S_*}\sup_{\tau\in[\tau_0,\tau_L)}\bigl|\beta_0^{(j)}(\tau)\bigr| \ge a\lambda_{n*}/n + \bigl(C_1\exp(C_2 s\tau_U) + L\cdot c^{-1}\bigr)\sqrt{s\log(p\vee n)/n} - (a-1)c_2\sqrt{s\log n/n},$$

$$\inf_{j\in S^*}\sup_{\tau\in[\tau_L,\tau_U]}\bigl|\beta_0^{(j)}(\tau)\bigr| \ge a\lambda_n^*/n + \bigl(C_1\exp(C_2 s\tau_U) + L\cdot c^{-1}\bigr)\sqrt{s\log(p\vee n)/n} - (a-1)c_2\sqrt{s\log n/n},$$

the same conditions required by high dimensional quantile regression for the complete data in Fan, Fan and Barut (2014). Thus, AL-HDCQR requires slightly stronger signal conditions than L-HDCQR to achieve both estimation consistency and model selection consistency. However, it does not demand the knowledge about unknown quantities to conduct hard-thresholding.

We next present the weak convergence of the proposed AL-HDCQR estimator.

THEOREM 4.3. *Under Conditions* (C1)–(C6) *and Assumption* 3.1, *if* $n^{-1/2}\times\lambda_{n*}\|\boldsymbol{\omega}_k\|_{2,S_*} = o(1)$, $\inf_{j\notin S_*}\lambda_{n*}\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$, *for all* $\tau_0 < \tau_k < \tau_L$, *and* $n^{-1/2}\lambda_n^*\|\boldsymbol{\omega}_k\|_{2,S^*} = o(1)$, $\inf_{j\notin S^*}\lambda_n^*\omega_k^{(j)}/\sqrt{sn\log(p\vee n)} \to \infty$, *for all* $\tau_L \le \tau_k \le \tau_U$, *then* $n^{1/2}(\hat{\boldsymbol{\beta}}_S(\tau) - \boldsymbol{\beta}_{0S}(\tau))$ *converges weakly to a mean zero Gaussian process for* $\tau\in[\tau_0,\tau_U]$.

The details of the Gaussian process in Theorem 4.3 are regulated to the supplemental article [Zheng, Peng and He (2018)]; see Section B. Theorem 4.3 requires stronger signal conditions than Propositions 4.1 and 4.2 so that $n^{-1/2}\lambda_{n*}\|\boldsymbol{\omega}_k\|_{2,S_*}$ and $n^{-1/2}\lambda_n^*\|\boldsymbol{\omega}_k\|_{2,S_*}$, the bias introduced by the penalties is asymptotically negligible. We can show that if $\inf_{j\in S_*}\sup_{\tau\in[\tau_0,\tau_L)}|\beta_0^{(j)}(\tau)| \geq a\lambda_{n*}/n + (C_1\exp(C_2 s\tau_U) + L\cdot c^{-1})\sqrt{s\log(p\vee n)/n}$ and

$$\inf_{j\in S^*}\sup_{\tau\in[\tau_L,\tau_U]}\left|\beta_0^{(j)}(\tau)\right| \geq a\lambda_n^*/n + \left(C_1\exp(C_2 s\tau_U) + L\cdot c^{-1}\right)\sqrt{s\log(p\vee n)/n},$$

the conditions in Theorem 4.3 can be satisfied by using the uniform weights and $\chi_2(\cdot)$. If we adopt the point-wise weights, the sufficient signal strength to satisfy conditions in Theorem 4.3 is $\inf_{j\in S_*}\inf_{\tau\in[\tau_0,\tau_L)}|\beta_0^{(j)}(\tau)| \geq a\lambda_{n*}/n + (C_1\exp(C_2 s\tau_U) + L\cdot c^{-1})\sqrt{s\log(p\vee n)/n}$ and $\inf_{j\in S^*}\inf_{\tau\in[\tau_L,\tau_U]}|\beta_0^{(j)}(\tau)| \geq a\lambda_n^*/n + (C_1\exp(C_2 s\tau_U) + L\cdot c^{-1})\sqrt{s\log(p\vee n)/n}$. Therefore, Theorem 4.3 entails weaker signal conditions with the choice of uniform weights than that from pointwise weighted penalty.

4.4. *The choice of tuning parameter.* It is known the performance of penalized methods greatly depends on the choice of tuning parameters. Because some unknown parameters are involved in conditions from Propositions 4.1 and 4.2, the theoretically optimal tuning parameters that satisfy those conditions cannot be prescribed in practice.

For uncensored high dimensional quantile regression, Zheng, Peng and He (2015) studied a GIC type tuning parameter selector, which is obtained as a minimizer of the following function:

$$\text{GIC}(\lambda;\Xi) := \int_{\Xi}\log\left(\mathbb{E}_n\rho_\tau\left(\log T_i - \mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau;\lambda)\right)\right)d\tau + \left|\hat{S}(\Xi;\lambda)\right|\phi_n, \tag{12}$$

where $\lambda$ is a candidate regularized parameter tuning the sparsity over $\Xi$, an interval of quantile levels; $\hat{\boldsymbol{\beta}}(\tau;\lambda)$ and $\hat{S}(\Xi;\lambda)$ are respectively a penalized quantile regression estimator and the corresponding selected model obtained with the tuning parameter $\lambda$ across $\tau\in\Xi$, and $\phi_n$ is a sequence converging to 0 with $n$. Zheng, Peng and He (2015) demonstrated that their GIC type selector can consistently identify the set of relevant variables over $\Xi$. In (12), $\int_\Xi\log(\mathbb{E}_n\rho_\tau(\log T_i - \mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau;\lambda)))\,d\tau$ serves to measure the overall model fitness over $\Xi$, which however is not obtainable in the presence of censoring, as some $T_i$'s are not observed.

To accommodate censoring, we propose to use $\hat{r}(\tau;\lambda) := \mathbb{E}_n[|\hat{D}_i(\hat{\boldsymbol{\beta}}(\tau;\lambda))|]$ to assess the fitness of model (1), where

$$\hat{D}_i\left(\hat{\boldsymbol{\beta}}(\tau;\lambda)\right) = \text{sign}\left(\hat{M}_i\left(\hat{\boldsymbol{\beta}}(\tau;\lambda)\right)\right)\sqrt{-2\left(\hat{M}_i\left(\hat{\boldsymbol{\beta}}(\tau;\lambda)\right)+\Delta_i\right)\log\left(\Delta_i - \hat{M}_i\left(\hat{\boldsymbol{\beta}}(\tau;\lambda)\right)\right)}$$

with $\hat{M}_i(\hat{\boldsymbol{\beta}}(\tau;\lambda)) = N_i(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau;\lambda)) - \int_{\tau_0}^{\tau} 1\{\log X_i \geq N_i(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(u;\lambda)\}\,dH(u) - \tau_0$. Here, $\hat{M}_i(\hat{\boldsymbol{\beta}}(\tau;\lambda))$ and $\hat{D}_i(\hat{\boldsymbol{\beta}}(\tau;\lambda))$, respectively, stand for the martingale residual and the deviance residual under the assumed model (1). Due to the skewness of martingale residuals, we choose to measure the overall model fitness based on deviance residuals (which is expected to produce a more normal shaped symmetric distribution through a transformation to the martingale residuals).

Our modified GIC type tuning parameter selector is defined as a minimizer of (13):

$$\text{GIC}(\lambda; \Xi) := \int_{\Xi} \log \hat{r}(\tau; \lambda)\, d\tau + \left|\hat{S}(\Xi; \lambda)\right| \phi_n. \tag{13}$$

By setting $\Xi$ as $[\tau_0, \tau_L)$ and $[\tau_L, \tau_U]$ sequentially in (13), we can get the practical tuning parameters $\hat{\lambda}_{n*}$ and $\hat{\lambda}_n^*$. Further investigations on the properties of this tuning parameter selector will require additional work at the technical level.

## 5. Numerical analysis.

5.1. *Simulation studies.* We conduct simulation studies to assess the finite sample performance of the proposed L-HDCQR and AL-HDCQR estimators. For the AL-HDCQR estimator, we consider the pointwise, uniform and average weights as described in Section 4. The proposed estimators are compared with Huang and Ma's (2010) variable selection method for the accelerated failure time (AFT) model, hereafter referred to as ALasso AFT method. We set the sample size $n = 300$ and the number of covariates $p = 400$. The quantile interval of interest is $[0.3, 0.7]$, that is, $\tau_L = 0.3$ and $\tau_U = 0.7$. For each simulation setup considered below, we set $\tau_0 = 0.1$ regardless of whether Assumption 3.1 is satisfied or not. The $\tau$-grid $\Gamma_{m_n}$ are $n/5$ equally space points between $\tau_0$ and $\tau_U$. In the L-HDCQR step, we first select a conservative $\tilde{\lambda}_{0,n}$ following Belloni and Chernozhukov (2011) or Fan, Fan and Barut (2014). For $k = 1, 2, \ldots, m_n$, we then use 5-fold cross-validation (CV) to select $\tilde{\lambda}_{k,n} - \tilde{\lambda}_{k-1,n}$, the increment of the tuning parameter, from a grid within the range, $(a\epsilon_n\sqrt{\log(p \vee n)n}, b\epsilon_n\sqrt{\log(p \vee n)n})$. For AL-HDCQR, we implement our GIC procedure with $\phi_n = \log(\log n) \log p/n$ and search the tuning parameter over a $\lambda$-grid in $[a'\sqrt{sn\log(p \vee n)\log\log n}, b'\sqrt{sn\log(p \vee n)\log\log n}]$. Here, $a$, $b$, $a'$ and $b'$ are pre-specified constants.

We consider the following five setups:

*Setup* (I): We generate the event times following the form

$$\log T_i = \tilde{\mathbf{Z}}_i^T \mathbf{b} + \varepsilon_i,$$

where the covariates $\tilde{\mathbf{Z}}_i$'s are generated from the multivariate normal distribution $N_p(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = (\sigma_{jk})_{p\times p}$ and $\sigma_{jk} = 0.5^{|j-k|}$ and truncated to be between $-5$ and 5, the coefficients $\mathbf{b}$ are set as $b^{(1)} = 2$, $b^{(2)} = 1.5$, $b^{(5)} = 4/5$, $b^{(10)} = 4/3$, $b^{(16)} = 1$, $b^{(25)} = 5/3$ and $b^{(j)} = 0$ for all other $j$'s, and $\varepsilon_i \sim N(0, 1)$, where $N(\cdot, \cdot)$ denotes the normal distribution. Therefore, $\boldsymbol{\beta}_0(\tau) = (Q_\varepsilon(\tau), \mathbf{b}^T)^T$ for all $\tau \in (0, 1)$ under model (1), where $Q_\varepsilon(\tau)$ denotes the $\tau$th quantile of the distribution of $\varepsilon$. The censoring time is generated as $\log C_i \sim N(0,16) + N(-6,1) + N(10,0.25)$, an equal probability mixture of $N(0,16)$, $N(-6,1)$, and $N(10,0.25)$, if $\tilde{Z}_i^{(1)} \geq 0$, and $N(0,16) + N(0,1)$, an equal probability mixture of $N(0,16)$ and $N(0,1)$, if $\tilde{Z}_i^{(1)} < 0$. In this case, the censoring distribution is covariate dependent. The censoring rate is around 35%.

*Setup* (II): The event times are generated following the same settings as in Setup (I) except $\varepsilon_i \sim \text{Laplace}(0, 3/2)$, the double exponential distribution with location parameter 0 and scale parameter 3/2. The censoring time is generated as $\log C_i \sim N(0,16) + N(-6,1) + N(10,0.25)$, which is covariate-independent. The censoring rate is around 25%.

*Setup* (III): The model used to generate the event times takes the form

$$\log T_i = \tilde{\mathbf{Z}}_i^T \mathbf{b} + 1.75 \tilde{Z}_i^{(3)} \varepsilon_i,$$

where $b^{(1)} = 2$, $b^{(10)} = 4/3$, $b^{(16)} = 4/5$, $b^{(25)} = 1$, and $b^{(j)} = 0$ for all other $j$'s, and $\varepsilon_i \sim N(0,1)$. We first generate $\acute{\mathbf{Z}}_i$ as in Setup (I), and then set $\tilde{\mathbf{Z}}_i = \acute{\mathbf{Z}}_i$, except $\tilde{\mathbf{Z}}_i^{(3)} = |\acute{\mathbf{Z}}_i^{(3)}| + 0.5$. Therefore, $\boldsymbol{\beta}_0(\tau)$ under (1) satisfies $\beta^{(1)}(\tau) = 0, \beta^{(4)}(\tau) = 1.75 Q_\varepsilon(\tau)$, and $\beta^{(j)}(\tau) = b^{(j+1)}$, for all other $j$'s. The censoring time is generated following the same distribution as in Setup (II). The censoring rate is round 30%.

*Setup* (IV): The event times are generated following

$$\log T_i = \tilde{\mathbf{Z}}_i^T \mathbf{b} + \phi_1(\xi_i) \tilde{Z}_i^{(1)} + \phi_5(\xi_i) \tilde{Z}_i^{(5)},$$

where $b^{(8)} = 2$, $b^{(15)} = 1$, $b^{(25)} = 1$ and $b^{(j)} = 0$ for all other $j$'s, $\phi_1(\cdot)$, $\phi_5(\cdot)$ are two functions plotted in Figure 1, and $\xi_i \sim \text{Unif}(0,1)$. We first generate $\acute{\mathbf{Z}}_i$ as in Setup (III), and then set $\tilde{\mathbf{Z}}_i^{(1)} = |\acute{\mathbf{Z}}_i^{(1)}|$, $\tilde{\mathbf{Z}}_i^{(5)} = |\acute{\mathbf{Z}}_i^{(5)}|$, and $\tilde{\mathbf{Z}}_i^{(j)} = \acute{\mathbf{Z}}_i^{(j)}$, for all other $j$'s. Then we can show that the quantile coefficient $\boldsymbol{\beta}_0(\tau)$ under (1) takes the form $\beta^{(1)}(\tau) = 0, \beta^{(2)}(\tau) = \phi_1(\tau), \beta^{(6)}(\tau) = \phi_5(\tau)$, and $\beta^{(j)}(\tau) = b^{(j+1)}$ for all other $j$'s. The censoring time is generated as $\log C_i \sim N(0,16) + N(-6,1) + N(12,0.36)$ if $\tilde{Z}^{(8)} \geq 0$ and $N(0,16) + N(0,1)$ if $\tilde{Z}^{(8)} < 0$. The censoring rate is around 25%.

*Setup* (V): The event times are generated following the same settings as in Setup (I) but we consider the fixed censoring time $\log C_i = 2$. The censoring rate is around 30%.

For each simulation setup, we conducted 200 replications. In each replication, the following three performance measures are calculated:

(1) Number of correctly identified relevant covariates over $[\tau_L, \tau_U]$, denoted by NC;
(2) Number of incorrectly selected covariates over $[\tau_L, \tau_U]$, denoted by NI;
(3) The relative absolute estimation errors with respect to the unpenalized CQR estimator $\hat{\boldsymbol{\beta}}^o(\tau)$ over $[\tau_L, \tau_U]$ provided the true model is used. It is defined as

$$\text{REE}_o = \frac{\int_{\tau_L}^{\tau_U} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\|_1 \, d\tau}{\int_{\tau_L}^{\tau_U} \|\hat{\boldsymbol{\beta}}^o(\tau) - \boldsymbol{\beta}_0(\tau)\|_1 \, d\tau}.$$

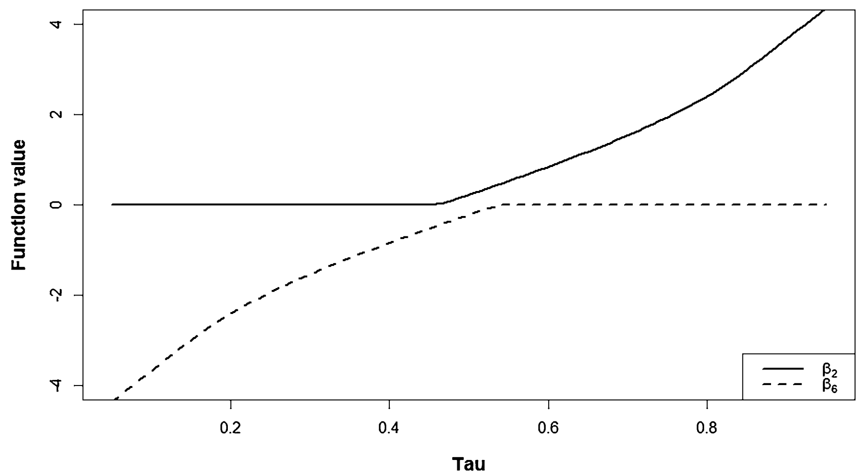


FIG. 1. *The quantile coefficient functions* $\beta_2(\cdot)$, $\beta_6(\cdot)$ *in simulation setup* (IV).

For ALasso AFT, we calculate $\mathrm{REE}_o$ by extrapolating the coefficient function estimate as a constant function over $[\tau_L, \tau_U]$.

The NC and NI evaluate the model selection performance. A good method is expected to produce NC and NI close to the true number of relevant covariates and 0, respectively. The criterion $\mathrm{REE}_o$ measures the global estimation accuracy over the quantile index interval of interest. In addition, we classify the selected model as under-fitted, correctly-fitted and over-fitted.

Table 1 presents the averages of NC, NI, the percentages of under-fitted, correctly-fitted and over- fitted models (PUF, PCF and POF), and the median of $\mathrm{REE}_o$ ($\mathrm{MREE}_o$) from 200 simulations.

In setup (I) and (II), all covariates have only constant quantile effects. The accelerated failure time assumption holds in both setups. It can be seen that the performance of ALasso AFT are quite different between setups (I) and (II). Given the errors follow a normal distribution in setup (I), it is reasonable to observe that ALasso AFT achieves a good PCF (95.0%) and the smallest $\mathrm{MREE}_o$ (0.847). However, in setup (II), where the errors become heavy-tailed, the model selection performance of ALasso AFT degrades considerably; the PUF is over 40% and the PCF is only around 35%. In contrast, the AL-HDCQR methods with the three different choices of weight (i.e., Pointwise, Average, Uniform) have very good performance in model selection in both setups with PCFs all above 85%. This suggests the proposed AL-HDCQR is robust against heavy-tailed errors, a desirable property inherited from quantile regression.

Setup (III) and (IV) are designed to assess the performance of different methods when some covariates have varying effects on different quantiles. In setup (III), the effect of covariate $\tilde{Z}^{(3)}$ at the $\tau$th conditional quantile of $\log T$ is $1.75\Phi^{-1}(\tau)$,

TABLE 1
*Simulation results of setup* (I)–(V)

| **Setup** | **Method** | **Weights** | **PUF(%)** | **PCF(%)** | **POF(%)** | **NC** | **NI** | **MREE$_o$** |
|---|---|---|---|---|---|---|---|---|
| Setup (I) | L-HDCQR | – | 0.0 | 0.0 | 100.0 | 6.000 | 112.945 | 6.906 |
| | | Pointwise | 1.0 | 98.0 | 1.0 | 5.990 | 0.010 | 2.538 |
| | AL-HDCQR | Average | 1.0 | 97.0 | 2.0 | 5.990 | 0.020 | 1.578 |
| | | Uniform | 0.5 | 98.5 | 1.0 | 5.995 | 0.010 | 1.541 |
| | ALasso AFT | – | 0.5 | 95.0 | 4.5 | 5.995 | 0.065 | 0.847 |
| Setup (II) | L-HDCQR | – | 0.0 | 0.0 | 100.0 | 6.000 | 147.810 | 9.447 |
| | | Pointwise | 5.5 | 88.5 | 6.0 | 5.945 | 0.065 | 1.521 |
| | AL-HDCQR | Average | 4.5 | 91.0 | 4.5 | 5.955 | 0.050 | 1.289 |
| | | Uniform | 2.0 | 93.0 | 5.0 | 5.980 | 0.055 | 1.255 |
| | ALasso AFT | – | 46.5 | 34.5 | 19.0 | 5.380 | 0.635 | 1.702 |
| Setup (III) | L-HDCQR | – | 0.5 | 0.0 | 99.5 | 4.995 | 158.435 | 9.827 |
| | | Pointwise | 11.0 | 82.5 | 6.5 | 4.887 | 0.871 | 1.569 |
| | AL-HDCQR | Average | 12.0 | 81.5 | 6.5 | 4.872 | 0.077 | 1.134 |
| | | Uniform | 11.0 | 81.5 | 7.5 | 4.882 | 0.097 | 1.140 |
| | ALasso AFT | – | 100.0 | 0.0 | 0.0 | 3.282 | 0.430 | 1.752 |
| Setup (IV) | L-HDCQR | – | 0.0 | 0.0 | 100.0 | 5.000 | 149.055 | 9.346 |
| | | Pointwise | 18.5 | 79.5 | 2.0 | 4.790 | 0.030 | 1.141 |
| | AL-HDCQR | Average | 24.5 | 75.5 | 0.0 | 4.755 | 0.005 | 1.200 |
| | | Uniform | 21.0 | 77.0 | 2.0 | 4.775 | 0.025 | 1.155 |
| | ALasso AFT | – | 100.0 | 0.0 | 0.0 | 2.955 | 0.350 | 2.236 |
| Setup (V) | L-HDCQR | – | 0.0 | 0.0 | 100.0 | 6.000 | 134.945 | 8.286 |
| | | Pointwise | 0.0 | 98.0 | 2.0 | 6.000 | 0.020 | 2.131 |
| | AL-HDCQR | Average | 0.0 | 98.5 | 1.5 | 6.000 | 0.015 | 1.977 |
| | | Uniform | 0.0 | 99.5 | 0.5 | 6.000 | 0.005 | 1.745 |
| | ALasso AFT | – | 2.0 | 96.5 | 1.5 | 5.980 | 0.005 | 0.879 |

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Therefore, the effect of $\tilde{Z}^{(3)}$ in this setup is 0 at $\tau = 0.5$ and becomes stronger as $\tau$ moves towards 0 or 1. Figure 1 presents the coefficient functions for $\tilde{Z}^{(1)}$ and $\tilde{Z}^{(5)}$ in setup (IV), showing that these two covariates have partial effects on two disjoint $\tau$-regions. Clearly, ALasso AFT suffers in these two cases. In both cases, its PUFs are 100%, and NCs are 3.282 and 2.955, respectively. These results indicate that ALasso AFT cannot consistently select relevant variables with varying effects. Contrariwise, the proposed AL-HDCQR methods can satisfactorily identify all relevant covariates over the quantile region of interest; the PCFs in setups (III) and (IV) are still above 75%.

Setup (V) is the same as setup (I) except that it has fixed censoring. Table 1 suggests that the proposed AL-HDCQR methods still perform well with good PCFs (close 100%) and small MREE$_o$s when the censoring variable is a constant. From Table 1, we also observe that L-HDCQR method tends to yield an over-fitted model with POF around 100% in all five setups. This is in line with our theoretical results in Corollary 4.2. Compared to L-HDCQR, the introduction of weighted penalties in AL-HDCQR significantly reduce the coefficient estimation bias as expected. Although the simulation results demonstrate that all three AL-HDCQR estimators are able to achieve consistency in both model selection and coefficient estimation, as proved in Propositions 4.1, 4.2 and Theorem 4.2, by comparing the results produced by three AL-HDCQR estimators, we find that the one with the uniform weight in general outperforms the estimator based on the pointwise or average weight. Relative to the pointwise weights, the uniform weights better reflect the global signal strength and hence reduce the variability of penalties across $[\tau_L, \tau_U]$. Compared with the average weights, the uniform weights enable better power for detecting the weak signals. Therefore, we recommend adopting the uniform weights in the practical application of the proposed AL-HDCQR procedure.

We also examined the robustness of our proposed estimators to the choice of $[\tau_L, \tau_U]$. We considered setup (IV) with $[\tau_L, \tau_U] = [0.29, 0.71]$ and $[0.31, 0.69]$, which are resulted from applying small perturbations to $[0.3, 0.7]$ and yet may represent the same interest in the conditional distribution as $[0.3, 0.7]$. It would be desirable to obtain similar variable selection and estimation results between these three choices of $[\tau_L, \tau_U]$. The detailed simulation study is reported in the supplementary material (see Section A of the supplemental article [Zheng, Peng and He (2018)]). The results suggest that our proposed estimators are robust to reasonable variations in the choice of $[\tau_L, \tau_U]$.

Additional simulations were conducted to evaluate the proposed methods in cases with heavy censoring or a heavy-tailed error distribution, and scenarios where the covariate sparsity at upper quantiles are of interest. Please see Section A of the supplemental article [Zheng, Peng and He (2018)] for more details.

5.2. *Real data analysis*. We now illustrate the proposed HDCQR estimators by the analysis of a real dataset. Our data comes from a large retrospective study [Shedden et al. (2008)] that used gene expression values to predict the survival time in lung cancer, the leading cause of cancer death in the United States. This microarray data set contains expression values of 22,283 genes and the survival time on 442 lung adenocarcinomas. The median follow-up time is 46 months. About 46% subjects had censored survival time.

Following Huang, Ma and Zhang (2008) and Wang, Wu and Li (2012), we first carry out data preprocessing: (step 1) remove observations with missing values; (step 2) exclude each gene whose maximum expression value among subjects in study was less than the 25th percentile of the entire expression values; (step 3) remove each gene that lacked sufficient variability. For a gene to be considered "sufficiently variable," we require the range of its expression values is no less than 2. There are 440 subjects and 15,983 genes left after these preprocessing steps. We next select 3000 genes with the largest variances. From these 3000 genes, we further choose the top 600 genes that have the largest correlation coefficients with the

TABLE 2
*Analysis results of lung cancer data over two* $[\tau_L, \tau_U]$

| $[\tau_L, \tau_U]$ | Method | Weights | All data # of genes selected | Random partition prediction error: mean (sd) |
|---|---|---|---|---|
| [0.1, 0.675] | AL-HDCQR | Pointwise | 11 | 0.855 (0.073) |
| | | Average | 10 | 0.854 (0.072) |
| | | Uniform | 11 | 0.840 (0.057) |
| | ALasso AFT | – | 4 | 1.195 (0.142) |
| [0.1, 0.700] | AL-HDCQR | Pointwise | 11 | 0.853 (0.073) |
| | | Average | 10 | 0.853 (0.074) |
| | | Uniform | 11 | 0.843 (0.059) |
| | ALasso AFT | – | 4 | 1.196 (0.128) |

observed log survival time. Then we apply the proposed methods and ALasso AFT to investigate the impact of these 600 genes on lung cancer survival time.

We apply the proposed methods with $[\tau_L, \tau_U] = [0.1, 0.675]$ and $[0.1, 0.700]$. Such choices of $[\tau_L, \tau_U]$ reflect our interest in finding prognostic gene expression signatures for moderate-risk lung cancer cases. The number of covariates selected by these methods using all 440 subjects are reported in Table 2. The proposed AL-HDCQR estimator with average weighted penalties selected 10 genes, which are included in the 11 genes found by AL-HDCQR estimators with pointwise or uniform weights. Our methods identify the same set of genes over the two different choices of intervals, suggesting the robustness of our method to small changes in the specification of $[\tau_L, \tau_U]$. On the other hand, ALasso AFT selects 4 genes, of which 2 genes are in common with those identified by AL-HDCQR methods.

To evaluate the AL-HDCQR methods, we also compute their prediction error as follows. We randomly split the 440 subjects into a training data set with 300 subjects and a testing data set with the other 140 subjects. We then apply the AL-HDCQR and ALasso AFT method into the training set and obtain the estimator $\hat{\boldsymbol{\beta}}(\tau)$. Next, we calculate the prediction errors over the quantile interval in the testing data set. For AL-HDCQR approaches, the prediction errors over $[\tau_L, \tau_U]$ are calculated as

$$\mathrm{PE}^{cqr}(\tau_L, \tau_U) = \frac{\sum_{i=1}^{n} 1\{i \text{ in testing set}\} \int_{\tau_L}^{\tau_U} |\hat{D}_i(\hat{\boldsymbol{\beta}}(\tau))| \, d\tau}{\sum_{i=1}^{n} 1\{i \text{ in testing set}\}},$$

where $\hat{D}_i(\hat{\boldsymbol{\beta}}(\tau))$ is the deviance residual from censored quantile regression defined in Section 4.4. For ALasso AFT, we extrapolate the estimated coefficient function as a constant function.

We present the averages of PE along with the corresponding standard deviations (within parentheses) based on 200 replications of random splitting into training and test sets in Table 2.

Table 2 shows that the proposed AL-HDCQR approaches selected more genes and producing much smaller prediction errors, as compared to the ALasso AFT method. The standard deviations of the prediction errors are also much smaller for the AL-HDCQR approaches. The ALasso AFT method producing fewer selected gene in this example is consistent with our simulation results, which suggest that the ALasso AFT method is highly likely to miss the relevant variables with varying covariate effects [e.g., PUF $= 100\%$ in setups (III) and (IV)]. Therefore, the less desirable performance of the ALasso AFT method in this example may result from the violation of the constant covariate effect assumption of the AFT model. It is also worth noting that the results from AL-HDCQR are impressively consistent among different choices of weights and $[\tau_L, \tau_U]$. This once again endorses the robustness of the proposed AL-HDCQR methods.

## APPENDIX A: PROPOSITIONS A.1 AND A.2

In this section, we present Propositions A.1 and A.2 and provide some discussions.

PROPOSITION A.1. *Under Conditions* (C1)–(C6) *and Assumption* 3.1, *one can find* $\tilde{\lambda}_{0,n}$ *of order* $\sqrt{\log(p \vee n)n}$ *and a sufficiently large constant* $C_1$ *such that the event* $\tilde{\Omega}_0(C_1, 0)$ *holds with probability at least* $1 - 16\exp(-4\log(p \vee n)) - 2\exp(-3\log(p \vee n))$.

Proposition A.1 lays the foundation to control the cumulative estimation errors. Next, we show that there exists a constant $C_2$, such that under event $\tilde{\Omega}_{k-1}(C_1, C_2)$, event $\tilde{\Omega}_k(C_1, C_2)$ holds with probability tending to 1.

PROPOSITION A.2. *Suppose Conditions* (C1)–(C6) *hold*, *there exists a universal constant* $C_2$ *such that under event* $\tilde{\Omega}_{k-1}(C_1, C_2)$, $1 \le k \le m_n$, *if*

$$\text{(a)}\quad 2c_0\Bigg[5\sqrt{\tilde{\tau}_k \log(p \vee n)n} + \frac{6\bar{f}C_1c_0}{c_0 - 1}\frac{\epsilon_n}{1-\tau_U}\sqrt{s}n\sum_{r=0}^{k-1}\nu_{r,n}(C_2) + \frac{8k\epsilon_n}{1-\tau_U}\log(p \vee n) + 3\sqrt{\log(p \vee n)n}\Bigg] \le \tilde{\lambda}_{k,n} \le c^{-1}\sqrt{\log(p \vee n)n},$$

$$\text{(b)}\quad \frac{6\bar{f}C_1c_0}{c_0 - 1}\frac{\epsilon_n}{1-\tau_U}\sqrt{s}n\nu_{k-1,n}(C_2) + \frac{8\epsilon_n}{1-\tau_U}\log(p \vee n) \le \tilde{\lambda}_{k,n} - \tilde{\lambda}_{k-1,n} \le c^{-1}\epsilon_n\sqrt{\log(p \vee n)n}$$

*and where* $\tilde{\tau}_k$ *is defined in Lemma* C.6 *and* $c$ *is some constant*, *then the event* $\tilde{\Omega}_k(C_1, C_2)$ *holds with probability at least* $1 - 4(5k + 7)\exp(-3\log(p \vee n))$.

Conditions (a) and (b) specify the theoretical conditions for tuning parameters $\tilde{\lambda}_{k,n}$'s. Condition (a) is similar to the strength conditions of tuning parameters that are widely imposed in high dimensional literature [Zhang and Huang (2008), Belloni and Chernozhukov (2011), for example]. The first inequality requires the penalization to be strong enough to produce the sparse estimates, while the second inequality provides an upper bound for penalization to avoid over-shrinkage. The item $6\bar{f}C_1c_0\epsilon_n\sqrt{s}n\sum_{r=0}^{k-1}\nu_{r,n}(C_2)/((c_0-1)(1-\tau_U))+8k\epsilon_n\log(p\vee n)/(1-\tau_U)$ in condition (a) is new, and serves as an upper bound for the cumulative estimation errors up to $\tau_{k-1}$. If one adopts a common tuning parameter for all $0\le k\le m_n$, say $\tilde{\lambda}_n$, then $C_1\ge 2c_0\tilde{\lambda}_n/(\lambda_{\min}(c_0-1)\sqrt{s\log(p\vee n)n})$ from the proof of Proposition A.1. Therefore, the cumulative errors can reach

$$
\begin{aligned}
&6\bar{f}C_1\frac{c_0}{(c_0-1)(1-\tau_U)}\epsilon_n\sqrt{s}n\sum_{r=0}^{k-1}\sqrt{s\log(p\vee n)/n}+\frac{8}{1-\tau_U}k\epsilon_n\log(p\vee n)\\
&\quad\ge 12\bar{f}\frac{c_0^2}{\lambda_{\min}(c_0-1)^2(1-\tau_U)}\sqrt{s}k\epsilon_n\tilde{\lambda}_n,
\end{aligned}
$$

which may exceed $\tilde{\lambda}_n$ as $k$ increases, resulting in insufficient penalization. Even if we choose a large $\tilde{\lambda}_n$, the insufficient penalization may still occur, as using a large tuning parameter at small $\tau_k$'s would lead to an unnecessarily large estimation error in $\tilde{\boldsymbol{\beta}}(\tau_k)$, and consequently the cumulative estimation errors. This indicates that adjusting $\tilde{\lambda}_{k,n}$ with $k$ is critical for achieving enough sparsity. By condition (b), one can increase the tuning parameter at each $k$ to offset the increase in the cumulative estimation error so that sparse estimates can still be realized. The increment of $\tilde{\lambda}_{k,n}$ satisfying condition (b) exists and is of asymptotic order $\epsilon_n\sqrt{\log(p\vee n)n}$. By condition (b), one can increase the tuning parameter at each $k$ to offset the increase in the cumulative estimation error so that sparse estimates can still be realized. The increment of $\tilde{\lambda}_{k,n}$ satisfying condition (b) exists and is of asymptotic order $\epsilon_n\sqrt{\log(p\vee n)n}$. With the modulation of $\tilde{\lambda}_{k,n}$, $\tilde{\Omega}_k(C_1,C_2)$ holds with probability tending to 1 and the estimation error of $\tilde{\boldsymbol{\beta}}(\tau_k)$ is bounded with the rate $C_1\nu_{k,n}$. It is worth noting that Lasso type penalties $\tilde{\lambda}_{k,n}$'s maintain the same form at different $k$, which facilitates the development of the properties of L-HDCQR.

## APPENDIX B: TECHNICAL PROOFS

We present the proofs of our main results in this section. All lemmas used in this section are provided in Appendix C.

PROOF OF PROPOSITION A.1. By Lemma C.2, we have with probability at least $1-2\exp(-3\log(p\vee n))$, $\tilde{\boldsymbol{\beta}}(\tau_0)-\boldsymbol{\beta}_0(\tau_0)\in A_{\tau_0}$. Therefore, we restrict our attention to event $\tilde{\Omega}_{0,0}:=\{\tilde{\boldsymbol{\beta}}(\tau_0)-\boldsymbol{\beta}_0(\tau_0)\in A_{\tau_0}\}$, and consider $\tilde{Q}_0(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\tilde{Q}_0(\boldsymbol{\beta}_0(\tau_0))$ for any $\boldsymbol{\delta}\in A_{\tau_0},\boldsymbol{\delta}^TE[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2$ and $t\le$

$\kappa/\sqrt{\lambda_{\min}}$. $\tilde{Q}_0(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\tilde{Q}_0(\boldsymbol{\beta}_0(\tau_0))$ can be decomposed as $\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))+\tilde{\lambda}_{0,n}(\|\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\|_1-\|\boldsymbol{\beta}_0(\tau_0)\|_1)$, where $\eta_{\tau_0}(\cdot)$ is defined in Lemma C.2. We first evaluate $\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))$. It can be written as $E[\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))]+\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau))-E[\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))]$. By Lemma C.3, uniformly for $\boldsymbol{\delta}\in A_{\tau_0}$ that satisfies $\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2, t\le\kappa/\sqrt{\lambda_{\min}}$,

$$n^{-1}E\big[\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\big)\big]-n^{-1}E\big[\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)\big)\big]\ge \underline{g}t^2-2At^3/(3q). \tag{14}$$

Since $\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau))-E[\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})-\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))]$ can be written as $2\mathbb{G}_n[\rho_{\tau_0}(\log X_i-\mathbf{Z}_i^T(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}))-\rho_{\tau_0}(\log X_i-\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau_0))]$, then we have

$$\begin{aligned}\sup_{\boldsymbol{\delta}\in A_{\tau_0},\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2}&\big|\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\big)-\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)\big)\\&-E\big[\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\big)-\eta_{\tau_0}\big(\boldsymbol{\beta}_0(\tau_0)\big)\big]\big|\\&\le 2\sqrt{n}\mathcal{A}_0(t),\end{aligned}$$

where $\mathcal{A}_0(t)$ is defined in Lemma C.4. According to Lemma C.4,

$$\begin{aligned}&\Pr\big(\mathcal{A}_0(t)\ge 48\sqrt{2}c_0\sqrt{s\log(p\vee n)}t/\big((c_0-1)\sqrt{\lambda_{\min}}\big)\big)\\&\quad\le 16p\exp\big(-4\log(p\vee n)\big).\end{aligned} \tag{15}$$

It is easy to see that

$$\begin{aligned}\sup_{\boldsymbol{\delta}\in A_{\tau_0},\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2}&\tilde{\lambda}_{0,n}\big|\|\boldsymbol{\beta}_0(\tau_k)+\boldsymbol{\delta}\|_1-\|\boldsymbol{\beta}_0(\tau_k)\|_1\big|\\&\le 2\tilde{\lambda}_{0,n}c_0\sqrt{s}t/\big((c_0-1)\sqrt{\lambda_{\min}}\big).\end{aligned} \tag{16}$$

If we choose $\tilde{\lambda}_{0,n}=8c_0\sqrt{\tau_0(1-\tau_0)\log(p\vee n)n}$, then by (14), (15) and (16), we obtain that

$$\begin{aligned}\inf_{\boldsymbol{\delta}\in A_{\tau_0},\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2}&n^{-1}\big[\tilde{Q}_0\big(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\big)-\tilde{Q}_0\big(\boldsymbol{\beta}_0(\tau_0)\big)\big]\\&\ge t\bigg\{\underline{g}t-\frac{2}{3q}At^2-96\sqrt{2}c_0\sqrt{s\log(p\vee n)/n}/\big((c_0-1)\sqrt{\lambda_{\min}}\big)\\&\quad-16c_0\sqrt{\tau_0(1-\tau_0)}c_0\sqrt{s\log(p\vee n)/n}/\big((c_0-1)\sqrt{\lambda_{\min}}\big)\bigg\}\end{aligned}$$

with probability at least $1-16\exp(-4\log(p\vee n))$. Therefore, there exists a sufficiently large constant $C_1$, such that

$$\inf_{\boldsymbol{\delta}\in A_{\tau_0},\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=\lambda_{\min}C_1^2 s\log(p\vee n)/n}\tilde{Q}_0\big(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}\big)-\tilde{Q}_0\big(\boldsymbol{\beta}_0(\tau_0)\big)>0.$$

Since (7) is convex with respect to $\mathbf{h}$, we have with probability at least $1 - 16\exp(-4\log(p\vee n)) - 2\exp(-3\log(p\vee n))$, $\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}(\tilde{\boldsymbol{\beta}}(\tau_0) - \boldsymbol{\beta}_0(\tau_0))\| \le \sqrt{\lambda_{\min}}C_1\sqrt{s\log(p\vee n)/n}$. By Condition (C5), $\|\tilde{\boldsymbol{\beta}}(\tau_0) - \boldsymbol{\beta}_0(\tau_0)\| \le C_1\sqrt{s\log(p\vee n)/n}$. This completes the proof of Proposition A.1. □

PROOF OF PROPOSITION A.2. We choose some constant $C_2$ such that

$$\begin{aligned} C_2 > {} & 4\frac{\bar{f}}{(1-\tau_U)s\underline{g}} + 64\sqrt{2}\frac{1}{(1-\tau_U)s\underline{g}C_1\lambda_{\min}}\frac{c_0}{c_0-1} \\ & + 8\frac{1}{(1-\tau_U)\underline{g}}\frac{c_0}{c_0-1}\frac{1}{\sqrt{s}\sqrt{\lambda_{\min}}} \\ & + 12\frac{\bar{f}}{(1-\tau_U)\underline{g}}\frac{c_0}{(c_0-1)^2}\frac{1}{\lambda_{\min}}, \end{aligned}$$

where $C_1$ is from Proposition A.1, $\underline{g}$ is defined in Condition (C2), and, $c_0$ and $\lambda_{\min}$ are defined in Condition (C5). It can be seen that the choice of $C_2$ does not depend on $n$. Then we show under the event $\tilde{\Omega}_{k-1}(C_1, C_2)$, event $\tilde{\Omega}_k(C_1, C_2)$ holds with large probability.

By Lemma C.10, if

$$\begin{aligned} \tilde{\lambda}_{k,n} \ge 2c_0\Bigg[ & 5\sqrt{\tilde{\tau}_k\log(p\vee n)n} + 6\bar{f}C_1\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{c_0-1}\sqrt{s}n\sum_{r=0}^{k-1}\nu_{r,n}(C_2) \\ & + \frac{\epsilon_n}{1-\tau_U}8k\log(p\vee n) + 3\sqrt{\log(p\vee n)n}\Bigg], \end{aligned}$$

then under $\tilde{\Omega}_{k-1}(C_1, C_2)$ given $1 \le k \le m_n$, with probability at least $1 - 4(k+1)\exp(-3\log(p\vee n))$, we have $\tilde{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}_0(\tau_k) \in A_{\tau_k}$. Therefore, we restrict our attention on $\tilde{\Omega}_{k,0} := \{\tilde{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}_0(\tau_k) \in A_{\tau_k}\}$.

We follow the similar arguments used in Proposition A.1. By Lemmas C.11, C.12 and C.13, we have under $\tilde{\Omega}_{k-1}(C_1, C_2)$,

$$\begin{aligned} & \inf_{\boldsymbol{\delta}\in A_{\tau_k}, \boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta} = t^2} n^{-1}[\tilde{Q}_k(\boldsymbol{\beta}_0(\tau_k) + \boldsymbol{\delta}) - \tilde{Q}_k(\boldsymbol{\beta}_0(\tau_k))] \\ & \quad \ge t\Bigg\{\underline{g}t - \frac{2}{3q}At^2 - 2\frac{\epsilon_n}{1-\tau_U}\sum_{r=0}^{k-1}(\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2) + L\epsilon_n) \\ & \qquad - 80\sqrt{2}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n} \\ & \qquad - 32\tau_0\sqrt{2}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n} \end{aligned}$$

$$- 64\sqrt{2}\sum_{k=0}^{k-1}\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n}$$

$$- 2\frac{\epsilon_n}{1-\tau_U}\sum_{k=0}^{k-1}\left(4C_1\frac{c_0}{c_0-1}\sqrt{s}\nu_{r,n}(C_2)+\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2)+L\epsilon_n\right)$$

$$- 2\sqrt{s}\frac{\tilde{\lambda}_{k,n}}{n}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\Bigg\}$$

with probability at least $1-4(5k+7)\exp(-3\log(p\vee n))$. Therefore, we have

$$\begin{aligned}
0 \le\ & \underline{g}C_1\sqrt{\lambda_{\min}}\nu_{k-1,n}(C_2)-\frac{2}{3q}A\lambda_{\min}\big(C_1\nu_{k-1,n}(C_2)\big)^2\\
& - 2\frac{\epsilon_n}{1-\tau_U}\sum_{r=0}^{k-2}\big(\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2)+L\epsilon_n\big)\\
& - 80\sqrt{2}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n}\\
& - 32\tau_0\sqrt{2}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n} \qquad (17)\\
& - 64\sqrt{2}\sum_{k=0}^{k-2}\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n}\\
& - 2\frac{\epsilon_n}{1-\tau_U}\sum_{k=0}^{k-2}\left(4C_1\frac{c_0}{c_0-1}\sqrt{s}\nu_{r,n}(C_2)+\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2)+L\epsilon_n\right)\\
& - 2\sqrt{s}\frac{\tilde{\lambda}_{k-1,n}}{n}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}},
\end{aligned}$$

under $\tilde{\Omega}_{k-1}(C_1, C_2)$.

Let $\nu_{k,n}(C_2)=(1+C_2 s\epsilon)\nu_{k-1,n}(C_2)$. If we choose $\tilde{\lambda}_{k,n}-\tilde{\lambda}_{k-1,n}=\frac{6\bar{f}C_1c_0}{c_0-1}\times\frac{\epsilon_n}{1-\tau_U}\sqrt{s}n\nu_{k-1,n}(C_2)+\frac{8\epsilon_n}{1-\tau_U}\log(p\vee n)$, then simple algebra yields that

$$\begin{aligned}
& \underline{g}C_1\sqrt{\lambda_{\min}}\nu_{k-1,n}(C_2)-\frac{2}{3q}A\lambda_{\min}\big(C_1(1+C_2 s\epsilon)\nu_{k-1,n}(C_2)\big)^2\\
& - 2\frac{\epsilon_n}{1-\tau_U}\sum_{r=0}^{k-2}\big(\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{k-1,n}(C_2)+L\epsilon_n\big)\\
& - 80\sqrt{2}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n}\\
& - 32\tau_0\sqrt{2}c_0\sqrt{s\log(p\vee n)/n}/\big((c_0-1)\sqrt{\lambda_{\min}}\big)
\end{aligned}$$

$$
\begin{aligned}
&- 64\sqrt{2}\sum_{k=0}^{k-2}\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n} \\
&- 2\frac{\epsilon_n}{1-\tau_U}\sum_{k=0}^{k-2}\biggl(4C_1\frac{c_0}{c_0-1}\sqrt{s}\nu_{r,n}(C_2)+\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2)+L\epsilon_n\biggr) \\
&- 2\sqrt{s}\frac{\tilde{\lambda}_{k-1,n}}{n}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}} \\
&+ \underline{g}C_1C_2\sqrt{\lambda_{\min}}s\epsilon_n\nu_{k-1,n}(C_2) - 2\frac{\epsilon_n}{1-\tau_U}\bigl(\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{k-1,n}(C_2)+L\epsilon_n\bigr) \\
&- 64\sqrt{2}\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\sqrt{s\log(p\vee n)/n} \\
&- 2\frac{\epsilon_n}{1-\tau_U}\biggl(4C_1\frac{c_0}{c_0-1}\sqrt{s}\nu_{k-1,n}(C_2)+\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{k-1,n}(C_2)+L\epsilon_n\biggr) \\
&- 2\sqrt{s}\frac{c_0}{(c_0-1)\sqrt{\lambda_{\min}}}\biggl(6\bar{f}C_1\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{c_0-1}\sqrt{s}\nu_{k-1,n}(C_2) \\
&+ 8\frac{\epsilon_n}{1-\tau_U}\frac{\log(p\vee n)}{n}\biggr) > 0
\end{aligned}
$$

by our choice of $C_2$. Again, since (7) is convex with respect to $\mathbf{h}$, under $\tilde{\Omega}_{k-1}(C_1, C_2)$, we have with probability at least $1-4(5k+7)\exp(-3\log(p\vee n))$,

$$
\bigl\|\bigl(E[\mathbf{Z}_i\mathbf{Z}_i^T]\bigr)^{1/2}\bigl(\tilde{\boldsymbol{\beta}}(\tau_k)-\boldsymbol{\beta}_0(\tau_k)\bigr)\bigr\| \le \sqrt{\lambda_{\min}}C_1\nu_{k,n}(C_2).
$$

By condition (C5), we have $\|\tilde{\boldsymbol{\beta}}(\tau_k)-\boldsymbol{\beta}_k(\tau_k)\| \le C_1\nu_{k,n}(C_2)$. This completes the proof of Proposition A.2. □

PROOF OF THEOREM 4.1. By Propositions A.1 and A.2, we have with probability at least $1-\sum_{r=0}^{m_n}4(5r+7)\exp(-3\log(p\vee n))$,

$$
\begin{aligned}
&\sup_{\tau_0\le\tau\le\tau_U}\|\tilde{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\| \\
&\quad\le \max\Bigl\{\sup_{\tau_k\le\tau<\tau_{k+1},k=0,\dots,m_n-1}\|\tilde{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\|, \|\tilde{\boldsymbol{\beta}}(\tau_{m_n})-\boldsymbol{\beta}_0(\tau_{m_n})\|\Bigr\} \\
&\quad\le \max\Bigl\{\max_{k=0,\dots,m_n-1}\|\tilde{\boldsymbol{\beta}}(\tau_k)-\boldsymbol{\beta}_0(\tau_k)\| \\
&\qquad + \sup_{\tau_k\le\tau<\tau_{k+1},k=0,\dots,m_n-1}\|\boldsymbol{\beta}_0(\tau)-\boldsymbol{\beta}_0(\tau_k)\|, C_1\nu_{m_n,n}(C_2)\Bigr\} \\
&\quad\le \max\Bigl\{\max_{k=0,\dots,m_n-1}C_1\nu_{k,n}(C_2)+L\sqrt{s}\epsilon_n, C_1\nu_{m_n,n}(C_2)\Bigr\} \qquad (18) \\
&\quad\le C_1(1+C_2s\epsilon_n)^{m_n}+L\sqrt{s}\epsilon_n
\end{aligned}
$$

$$\le C_1(1+C_2 s\epsilon_n)^{\tau_U/\epsilon_n} + L\sqrt{s}\epsilon_n$$
$$\le C_1 \exp(C_2 s\tau_U)\sqrt{s\log(p\vee n)/n} + L\sqrt{s}c^{-1}\sqrt{\log(p\wedge n)/n}$$
$$\le \big(C_1\exp(C_2 s\tau_U) + L\cdot c^{-1}\big)\sqrt{s\log(p\vee n)/n},$$

where the third and sixth inequalities follow from Conditions (C4) and (C6). Since we can always find some constant $C_3$ satisfying $1 - C_3(p\vee n)^{-1} \le 1 - \sum_{r=0}^{m_n} 4(5r+7)\exp(-3\log(p\vee n))$ and $\epsilon_n \sim O(\sqrt{\log n/n})$, $\tilde{\boldsymbol{\beta}}(\tau)$ is uniformly consistent to $\boldsymbol{\beta}_0(\tau)$ with the convergence rate $\sqrt{s\log(p\vee n)/n}$ across $\tau\in[\tau_0,\tau_U]$. □

PROOF OF PROPOSITION 4.1. The proof of Proposition 4.1 follows the lines in Theorem A.1 and Theorems 3.1–3.3 in Zheng, Peng and He (2015). We can show there exists a constant $C_4$ such that

$$\big\|\big(E[\mathbf{Z}_i\mathbf{Z}_i^T]\big)^{1/2}\big(\hat{\boldsymbol{\beta}}(\tau_0) - \boldsymbol{\beta}_0(\tau_0)\big)\big\| \le \sqrt{\lambda_{\min}}C_4\sqrt{s\log n/n},\ \|\hat{\boldsymbol{\beta}}(\tau_0)-\boldsymbol{\beta}_0(\tau_0)\| \le C_4\sqrt{s\log n/n},$$

and $\{j : \hat{\boldsymbol{\beta}}^{(j)}(\tau_0)\neq 0\}\subset S_*$ with probability at least $1 - 38\exp(-3\log(p\vee n))$. We omit the details here. □

PROOF OF PROPOSITION 4.2. Let $C_5$ be some positive constant such that

$$C_5 > 4\frac{\bar{f}}{(1-\tau_U)\underline{g}s} + 32\sqrt{2}\frac{1}{(1-\tau_U)\lambda_{\min}\underline{g}sC_4} + 8\frac{1}{(1-\tau_U)\sqrt{s}\sqrt{\lambda_{\min}}\underline{g}}\frac{c_0}{c_0-1}.$$

We only show the case for all $\tau_L\le\tau_k\le\tau_U$. The other case can be proved with the same arguments. By Lemma C.14, if $\hat{\lambda}^*\|\boldsymbol{\omega}_k\|_{2,S^*}\le c_2\sqrt{sn\log n}$, then under event $\hat{\Omega}_{k-1}(C_4,C_5)$, we have $\|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}(\hat{\boldsymbol{\beta}}^o(\tau_k)-\boldsymbol{\beta}_0(\tau_k))\|\le\sqrt{\lambda_{\min}}C_4\vartheta_{k,n}(C_5)$, $\|\hat{\boldsymbol{\beta}}^o(\tau_k)-\boldsymbol{\beta}_0(\tau_k)\|\le C_4\vartheta_{k,n}(C_5)$, with probability at least $1-8(2k+3)\times\exp(-3\log(p\vee n))$, where $\hat{\boldsymbol{\beta}}^o(\tau)$ denotes the oracle penalized estimator. Since $\inf_{j\notin S^*}\hat{\lambda}^*\omega_k^{(j)}/\sqrt{n\log(p\vee n)}\to\infty$, then according to Lemma C.15, $\hat{\boldsymbol{\beta}}^o(\tau_k) = \hat{\boldsymbol{\beta}}(\tau_k)$ with probability at least $1-2(19k+22)\exp(-3\log p)$. This immediately implies that under event $\hat{\Omega}_{k-1}(C_4,C_5)$, $\hat{\Omega}_k(C_4,C_5)$ holds with probability at least $1-2(19k+22)\exp(-3\log p)$. □

PROOF OF THEOREM 4.2. The proof of Theorem 4.2 follows the lines in Theorem 4.1. We omit the details here. □

PROOF OF THEOREM 4.3. By Lemma C.16 and our choice of $\Gamma_{m_n}$, we have

$$\sup_{\tau_0\le\tau\le\tau_U}\sup_{j\in S}\big|\mathbb{G}_n\big[\mathbf{Z}_{iS}N_i\big(\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau)\big) - \mathbf{Z}_{iS}N_i\big(\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau)\big)\big]\big| \xrightarrow{p} 0. \tag{19}$$

By Lemma C.17, we also obtain

$$\begin{aligned}
&\sup_{\tau_0\le\tau\le\tau_U}\sup_{j\in S}\bigg|\mathbb{G}_n\bigg[\mathbf{Z}_{iS}\bigg(\int_{\tau_0}^{\tau} 1\{\log X_i \ge \mathbf{Z}_i^T\hat{\boldsymbol\beta}(u)\}\,dH(u)+\tau_0\bigg) \\
&\qquad - \bigg(\int_{\tau_0}^{\tau} 1\{\log X_i \ge \mathbf{Z}_i^T\boldsymbol\beta_0(u)\}\,dH(u)+\tau_0\bigg)\bigg]\bigg| \xrightarrow{p} 0. \qquad (20)
\end{aligned}$$

From (19) and (20), we obtain that

$$\begin{aligned}
&-n^{1/2}\mathbb{E}_n\bigg[\mathbf{Z}_{iS}\bigg(N_i(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau)) - \int_{\tau_0}^{\tau} 1\{\log X_i \ge \mathbf{Z}_i^T\boldsymbol\beta_0(u)\}\,dH(u)-\tau_0\bigg)\bigg] \\
&\quad = n^{1/2}\big(E\big[\mathbf{Z}_{iS}\big(N_i(\mathbf{Z}_i^T\hat{\boldsymbol\beta}(\tau)) - N_i(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau))\big)\big]\big) + n^{-1/2}\sup_{1\le k\le m_n}\hat\lambda_*\boldsymbol\omega_k \\
&\qquad - n^{1/2}\int_{\tau_0}^{\tau} E\big[\mathbf{Z}_{iS}1\{\log X_i \ge \mathbf{Z}_i^T\hat{\boldsymbol\beta}(u)\} \\
&\qquad - \mathbf{Z}_{iS}1\{\log X_i \ge \mathbf{Z}_i^T\boldsymbol\beta_0(u)\}\big]\,dH(u) + o_p(1) \\
&\quad = n^{1/2}\big(E\big[\mathbf{Z}_{iS}\big(N_i(\mathbf{Z}_i^T\hat{\boldsymbol\beta}(\tau)) - N_i(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau))\big)\big]\big) \\
&\qquad - \int_{\tau_0}^{\tau}\big(E\big[\mathbf{Z}_{i,S}\mathbf{Z}_{i,S}^T g(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau))\big]\big(E\big[\mathbf{Z}_{i,S}\mathbf{Z}_{i,S}^T f(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau))\big]\big)^{-1} + o_p(1)\big) \\
&\qquad \times n^{1/2}\big(E\big[\mathbf{Z}_{iS}\big(N_i(\mathbf{Z}_i^T\hat{\boldsymbol\beta}(\tau)) - N_i(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau))\big)\big]\big)\,dH(\tau) + o_p(1).
\end{aligned}$$

The rest of the proof follows from the same arguments used for Theorem 2 in Peng and Huang (2008). □

## APPENDIX C: LEMMAS

We present the technical lemmas used in the proofs of our theorems and propositions. The proofs of lemmas are relegated to the supplementary material [see Section C of Zheng, Peng and He (2018)].

LEMMA C.1. *Let* $\phi_i(\tau_0) = 2\tau_0 - 2\Delta_i 1\{\log X_i \le \mathbf{Z}_i^T\boldsymbol\beta_0(\tau_0)\}$. *Under Assumption* 3.1,

$$\Pr\Big(\sup_{1\le j\le p}\mathbb{E}_n\big[Z_i^{(j)}\phi_i(\tau_0)\big] > 8\sqrt{\tau_0(1-\tau_0)\log(p\vee n)/n}\Big) \le 2\exp\big(-3\log(p\vee n)\big),$$

*when n is sufficiently large.*

LEMMA C.2. *Under the same conditions from Lemma* C.1, *if* $\tilde\lambda_{0,n} \ge 8c_0\sqrt{\tau_0(1-\tau_0)\log(p\vee n)n}$, *then with probability at least* $1-2\exp(-3\log(p\vee n))$, $\tilde{\boldsymbol\beta}(\tau_0)-\boldsymbol\beta_0(\tau_0)\in A_{\tau_0}$.

LEMMA C.3. *Under Conditions* (C2), (C5) *and Assumption* 3.1, *given any* $0 < t \le \kappa/\sqrt{\lambda_{\min}}$, *we have* $n^{-1}E[\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta})] - n^{-1}E[\eta_{\tau_0}(\boldsymbol{\beta}_0(\tau_0))] \ge \underline{g}t^2 - 2At^3/(3q)$, *uniformly for* $\boldsymbol{\delta} \in A_{\tau_0}$ *that satisfies* $\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta} = t^2$.

LEMMA C.4. *Let* $\bar{\rho}_{\tau_0,i}(\mathbf{h}) := \Delta_i \rho_{\tau_0}(\log X_i - \mathbf{Z}_i^T\mathbf{h}) + \tau_0(1-\Delta_i)(\log X_i - \mathbf{Z}_i^T\mathbf{h})$ *and*

$$\mathcal{A}_0(t) := \sup_{\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta} \le t^2, \boldsymbol{\delta}\in A_{\tau_0}} \left|\mathbb{G}_n\left[\bar{\rho}_{\tau_0,i}(\boldsymbol{\beta}_0(\tau_0)+\boldsymbol{\delta}) - \bar{\rho}_{\tau_0,i}(\boldsymbol{\beta}_0(\tau_0))\right]\right|,$$

*under Conditions* (C1)–(C6) *and Assumption* 3.1, *we have*

$$\Pr(\mathcal{A}_0(t) \ge 12K_1) \le 16p\exp\left(-\frac{K_1^2}{2(2c_0\sqrt{s}t/((c_0-1)\sqrt{\lambda_{\min}}))^2}\right)$$

*for* $K_1 > t$.

LEMMA C.5. $\|\tilde{\boldsymbol{\beta}}(\tau_k)\|_0 \le n \wedge p$ *uniformly over* $1 \le k \le m$.

LEMMA C.6. *Let* $2\tilde{\tau}_k = 2\tau_k + 2H^2(\tau_k)$. *Suppose Condition* (C1) *holds*, *we have for all* $1 \le k \le m$

$$\Pr\left(\sup_{1\le j\le p}\left|\sum_{i=1}^n Z_i^{(j)}\left\{N_i(\mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau_k)) - \int_0^{\tau_k} 1\{\log X_i > \mathbf{Z}_i^T\boldsymbol{\beta}_0(u)\}\,dH(u)\right\}\right| > 5\sqrt{\tilde{\tau}_k\log(p\vee n)n}\right) \le 2\exp(-4\log(p\vee n)).$$

LEMMA C.7. *Suppose Condition* (C4) *holds*, *we have*, *for sufficiently large* $n$

$$\Pr\left(\sup_{1\le j\le p}\sum_{i=1}^n |Z_i^{(j)}|\left(1\{\log X_i > \mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau_k)\} - 1\{\log X_i > \mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau_{k+1})\}\right) > 4\log(p\vee n)\right) \le 2\exp(-3\log(p\vee n)).$$

LEMMA C.8. *Given* $0 \le k \le m-1$, *under conditions* (C1) *and* (C2), *if* $\|\tilde{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}_0(\tau_k)\| \le C_1\nu_{k,n}(C_2)$ *and* $\tilde{\boldsymbol{\beta}}(\tau_k) \in A_{\tau_k}$, *then for sufficiently large* $n$,

$$\Pr\left(\sup_{1\le j\le p}\left|\mathbb{E}_n\left[Z_i^{(j)}\left(1\{\log X_i - \mathbf{Z}_i^T\tilde{\boldsymbol{\beta}}(\tau_k) > 0\} - 1\{\log X_i - \mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau_k) > 0\}\right)\right]\right| > 6\bar{f}C_1c_0/(c_0-1)\sqrt{s}\nu_{k,n}(C_2)\right) \le 2\exp(-3\log(p\vee n)).$$

Let

$$\tilde{\phi}_{i,k}(\mathbf{u}) = N_i(\mathbf{Z}_i^T\mathbf{u}) - \left(\sum_{r=0}^{k-1}\int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i \geq \mathbf{Z}_i^T\tilde{\boldsymbol{\beta}}(\tau_k)\}\,dH(\tau) + \tau_0\right).$$

LEMMA C.9. *Suppose conditions* (C1)–(C6) *hold, for any* $1 \leq k \leq m_n$, *under event* $\tilde{\Omega}_{k-1}(C_1, C_2)$,

$$\begin{aligned}\Pr\biggl(n \max_{1\leq j\leq p}\bigl|\mathbb{E}_n\bigl[Z_i^{(j)}\tilde{\phi}_{i,k}(\boldsymbol{\beta}_0(\tau_k))\bigr]\bigr| &> 5\sqrt{\tilde{\tau}_k\log(p\vee n)n}\\ &+ 6\bar{f}C_1\frac{\epsilon_n}{1-\tau_U}\frac{c_0}{c_0-1}\sqrt{s}n\sum_{r=0}^{k-1}\nu_{r,n}(C_2)\\ &+ \frac{\epsilon_n}{1-\tau_U}8k\log(p\vee n) + 3\sqrt{\log(p\vee n)n}\biggr)\\ &\leq 4(k+1)\exp\bigl(-3\log(p\vee n)\bigr).\end{aligned}$$

LEMMA C.10. *Under conditions* (C1)–(C6), *under* $\tilde{\Omega}_{k-1}(C_1, C_2)$ *given* $1 \leq k \leq m$, *if*

$$\begin{aligned}\tilde{\lambda}_{k,n} \geq 2c_0\biggl[&5\sqrt{\tilde{\tau}_k\log(p\vee n)n} + 6\bar{f}C_1\frac{\epsilon}{1-\tau_U}\frac{c_0}{c_0-1}\sqrt{s}n\sum_{r=0}^{k-1}\nu_{r,n}(C_2)\\ &+ \frac{\epsilon}{1-\tau_U}8k\log(p\vee n) + 3\sqrt{\log(p\vee n)n}\biggr],\end{aligned}$$

*then with probability at least* $1-4(k+1)\exp(-3\log(p\vee n))$, $\tilde{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}_0(\tau_k) \in A_{\tau_k}$.

LEMMA C.11. *Under event* $\tilde{\Omega}_{k-1}(C_1, C_2)$, *given* $1 \leq k \leq m_n$, *if* $t \leq \kappa/\sqrt{\lambda_{\min}}$

$$\begin{aligned}&n^{-1}E\bigl[\tilde{L}_k\bigl(\boldsymbol{\beta}_0(\tau_k)+\boldsymbol{\delta}\bigr)\bigr] - n^{-1}E\bigl[\tilde{L}_k\bigl(\boldsymbol{\beta}_0(\tau_k)\bigr)\bigr]\\ &\quad\geq \underline{g}t^2 - \frac{2}{3q}At^3 - 2\frac{\epsilon_n}{1-\tau_U}\sum_{r=0}^{k-1}\bigl(\bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2) + L\epsilon_n\bigr)t\end{aligned}$$

*uniformly for* $\boldsymbol{\delta} \in A_{\tau_k}$ *satisfying* $\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta} = t^2$.

LEMMA C.12. *Given* $1 \leq k \leq m_n$, *let*

$$\begin{aligned}\mathcal{A}_k(t) := \sup_{\boldsymbol{\delta}\in A_{\tau_k},\boldsymbol{\delta}^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\boldsymbol{\delta}=t^2}&\bigl|\mathbb{G}_n\bigl[\Delta_i\bigl(\bigl|\log X_i - \mathbf{Z}_i^T(\boldsymbol{\beta}_0(\tau)+\boldsymbol{\delta})\bigr|\\ &- \bigl|\log X_i - \mathbf{Z}_i^T\boldsymbol{\beta}_0(\tau)\bigr|\bigr) + \Delta_i\mathbf{Z}_i^T\boldsymbol{\delta}\bigr]\bigr|,\end{aligned}$$

*under conditions* (C1)–(C6), *we have*

$$\Pr\big(\mathcal{A}_k(t) \geq 20K_1\big) \leq 16p\exp\bigg(-\frac{K_1^2}{2(2c_0\sqrt{s}t/((c_0-1)\sqrt{\lambda_{\min}}))^2}\bigg),$$

*for* $K_1 > t$.

LEMMA C.13. *Given* $1 \leq k \leq m_n$, *let*

$$\mathcal{C}_k(t) := \sup_{\delta\in A_{\tau_k}, \delta^T E[\mathbf{Z}_i\mathbf{Z}_i^T]\delta\leq t^2}\bigg|\mathbb{G}_n\bigg[\boldsymbol{\delta}^T\mathbf{Z}_i\bigg(\sum_{r=0}^{k-1}\int_{\tau_r}^{\tau_{r+1}} 1\{\log X_i > \mathbf{Z}_i^T\tilde{\boldsymbol{\beta}}(\tau_r)\}\,dH(u) + \tau_0\bigg)\bigg]\bigg|,$$

*under event* $\tilde{\Omega}_{k-1}(C_1, C_2)$, *we have*

$$\begin{aligned}\Pr\bigg(\mathcal{C}_k(t) > \sum_{r=0}^{k-1}\bigg[8\frac{\epsilon_n}{1-\tau_U}K_1 &+ \frac{\epsilon_n}{1-\tau_U}\bigg(4C_1\frac{c_0}{c_0-1}\sqrt{s}\nu_{r,n}(C_2) \\ &+ \bar{f}C_1\sqrt{\lambda_{\min}}\nu_{r,n}(C_2) + L\epsilon_n\bigg)\sqrt{n}t\bigg] + 4\tau_0K_1\bigg) \\ &\leq 8(2k+1)p\exp\bigg(-\frac{K_1^2}{2(2c_0\sqrt{s}t/((c_0-1)\sqrt{\lambda_{\min}}))^2}\bigg)\end{aligned}$$

*for any* $K_1 > t$.

LEMMA C.14. *Suppose conditions* (C1)–(C6) *hold, if* $\hat{\lambda}_*\|\boldsymbol{\omega}_k\|_{2,S_*} \leq c_2\sqrt{sn\log n}$ *for* $\tau_0 < \tau_k < \tau_L$, *and* $\hat{\lambda}^*\|\boldsymbol{\omega}_k\|_{2,S^*} \leq c_2\sqrt{sn\log n}$ *for* $\tau_L \leq \tau_k \leq \tau_U$, *then under event* $\hat{\Omega}_{k-1}(C_4, C_5)$, *for any* $1 \leq k \leq m_n$,

$$\big\|\big(E\big[\mathbf{Z}_i\mathbf{Z}_i^T\big]\big)^{1/2}\big(\hat{\boldsymbol{\beta}}^o(\tau_k) - \boldsymbol{\beta}_0(\tau_k)\big)\big\| \leq \sqrt{\lambda_{\min}}C_4\vartheta_{k,n}(C_5), \ \big\|\hat{\boldsymbol{\beta}}^o(\tau_k) - \boldsymbol{\beta}_0(\tau_k)\big\| \leq C_4\vartheta_{k,n}(C_5),$$

*with probability at least* $1 - 8(2k+3)\exp(-3\log n)$.

LEMMA C.15. *Suppose the regularity conditions* (C1)–(C6) *hold. Under*, $\hat{\Omega}_{k-1}(C_4, C_5)$, *if* $\inf_{j\notin S_*}\hat{\lambda}_*\omega_k^{(j)}/\sqrt{n\log p} \to \infty$, *for all* $\tau_0 < \tau_k < \tau_L$, *and*

$\inf_{j\notin S^*}\hat{\lambda}^*\omega_k^{(j)}/\sqrt{n\log p}\to\infty$ *for all* $\tau_L\le\tau_k\le\tau_U$, *such that*

$$\begin{aligned} c_3\sqrt{n\log p} :> \max_{k\le m_n}\Bigg\{&\bar{g}\sqrt{\lambda_{\min}}C_4\vartheta_{k,n}(C_5)n \\ &+20\lambda_{\min}^{1/4}\big(\bar{g}\lambda_{\max}C_4\vartheta_{k,n}(C_5)sn\log p/q\big)^{1/2} \\ &+n\frac{\epsilon_n}{1-\tau_U}\sum_{r=0}^{k-1}\big(L\epsilon_n+\bar{f}\sqrt{\lambda_{\min}}C_4\vartheta_{r,n}(C_5)\big) \\ &+3\sqrt{n\log p}-3\log(1-\tau_U)\sqrt{n\log p} \\ &+\sum_{r=0}^{k-1}\bigg(20\frac{\epsilon_n}{1-\tau_U}\lambda_{\min}^{1/4}\big(\bar{f}\lambda_{\max}C_4\vartheta_{r,n}(C_5)sn\log p/q\big)^{1/2} \\ &+4\frac{\epsilon_n}{1-\tau_U}\sqrt{\lambda_{\max}L\epsilon_n n}\bigg)\Bigg\} \end{aligned}$$

*and* $\hat{\boldsymbol\beta}^o(\tau_k)-\boldsymbol\beta_0(\tau_k)\in R_{S^*}(\tau_k)$, *where* $R_{S^*}(\tau_k)$ *denotes the restricted space* $\{\boldsymbol\delta:\|\boldsymbol\delta\|_{0,S^{*c}}=0, \|\boldsymbol\delta\|\le C_4\vartheta_{k,n}(C_5), \|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}\boldsymbol\delta\|\le\sqrt{\lambda_{\min}}C_4\vartheta_{k,n}(C_5)\}$, *then* $\hat{\boldsymbol\beta}^o(\tau_k)=\hat{\boldsymbol\beta}(\tau_k)$ *with probability at least* $1-2(9k+10)\exp(-3\log p)$.

LEMMA C.16. *Given any* $\boldsymbol\alpha\in\mathbb{R}^p$, *such that* (1) $\|\boldsymbol\alpha\|_{0,S^c}=0$ *and* $\|\boldsymbol\alpha\|=1$ *or* (2) $\boldsymbol\alpha=\mathbf{e}_j$, *where* $\mathbf{e}_j$ *is the* $j$*th column of the identity matrix, if* $\boldsymbol\delta\in R_S(\tau_k):=\{\boldsymbol\delta:\|\boldsymbol\delta\|_{0,S^c}=0, \|\boldsymbol\delta\|\le C_4\vartheta_{k,n}(C_5), \|(E[\mathbf{Z}_i\mathbf{Z}_i^T])^{1/2}\boldsymbol\delta\|\le\sqrt{\lambda_{\min}}C_4\vartheta_{k,n}(C_5)\}$, *then*

$$\begin{aligned} \Pr\Big(&\sup_{\boldsymbol\delta\in R_S(\tau_k)}\big|\mathbb{G}_n\big[\boldsymbol\alpha^T\mathbf{Z}_iN_i\big(\mathbf{Z}_i^T\big(\boldsymbol\beta_0(\tau_k)+\boldsymbol\delta\big)\big)-\boldsymbol\alpha^T\mathbf{Z}_iN_i\big(\mathbf{Z}_i^T\boldsymbol\beta_0(\tau_k)\big)\big]\big| \\ &\ge 20\lambda_{\min}^{1/4}\big(\bar{g}\lambda_{\max}C_4\vartheta_{k,n}(C_5)s\log n/q\big)^{1/2}\Big) \\ &\quad\le 16\exp(-4s\log n). \end{aligned}$$

LEMMA C.17. *Let* $\boldsymbol\alpha$ *be the same as in Lemma* C.16. *Under* $\hat{\Omega}_{k-1}(C_4,C_5)$, *we have*

$$\begin{aligned} \Pr\Bigg(&\Bigg|\mathbb{G}_n\Bigg[\boldsymbol\alpha^T\mathbf{Z}_i\Bigg(\sum_{r=0}^{k-1}\int_{\tau_r}^{\tau_{r+1}}1\{\log X_i\ge\mathbf{Z}_i^T\hat{\boldsymbol\beta}(\tau_r)\}\,dH(\tau) \\ &-\int_0^{\tau_k}1\{\log X_i\ge\mathbf{Z}_i^T\boldsymbol\beta_0(\tau)\}\,dH(\tau)\Bigg)\Bigg]\Bigg| \\ &\ge\sum_{r=0}^{k-1}20\frac{\epsilon_n}{1-\tau_U}\lambda_{\min}^{1/4}\big(\bar{f}\lambda_{\max}C_4\vartheta_{r,n}(C_5)s\log n/q\big)^{1/2}+4\frac{\epsilon_n}{1-\tau_U}\sqrt{\lambda_{\max}L\epsilon_n}\Bigg) \\ &\le 18k\exp(-4\log n). \end{aligned}$$

**Acknowledgments.** The authors are grateful for helpful comments from an anonymous Associate Editor and referees.

## SUPPLEMENTARY MATERIAL

**Supplement to "High dimensional censored quantile regression"** (DOI: 10.1214/17-AOS1551SUPP; .pdf). Additional simulation results, remarks, and proofs of technical lemmas.

## REFERENCES


BELLONI, A. and CHERNOZHUKOV, V. (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. MR3012402

EFRON, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability IV* 831–853. Prentice Hall, Engewood Cliffs, NJ.

FAN, J., FAN, Y. and BARUT, E. (2014). Adaptive robust variable selection. *Ann. Statist.* **42** 324–351. MR3189488

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York. MR1100924

HE, X. and SHAO, Q.-M. (2000). On parameters of increasing dimensions. *J. Multivariate Anal.* **73** 120–135. MR1766124

HUANG, J. and MA, S. (2010). Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Anal.* **16** 176–195. MR2608284

HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. MR2247210

HUANG, J., MA, S. and ZHANG, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statist. Sinica* **18** 1603–1618. MR2469326

JOHNSON, B. A. (2009). On lasso for censored data. *Electron. J. Stat.* **3** 485–506. MR2507457

KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge.

KOENKER, R. (2016). quantreg: Quantile regression. R package version 5.29.

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

NEOCLEOUS, T., BRANDEN, K. V. and PORTNOY, S. (2006). Correction to: "Censored regression quantiles" [*J. Amer. Statist. Assoc.* **98** (2003) 1001–1012; MR2041488] by Portnoy. *J. Amer. Statist. Assoc.* **101** 860–861. MR2281250

PENG, L. (2012). Self-consistent estimation of censored quantile regression. *J. Multivariate Anal.* **105** 368–379. MR2877523

PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103** 637–649. MR2435468

PENG, L., XU, J. and KUTNER, N. (2014). Shrinkage estimation of varying covariate effects based on quantile regression. *Stat. Comput.* **24** 853–869. MR3229701

PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. MR2041488

PORTNOY, S. and LIN, G. (2010). Asymptotics for censored regression quantiles. *J. Nonparametr. Stat.* **22** 115–130. MR2598957

POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. MR0853049

SHEDDEN, K., TAYLOR, J. M. et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nature Medicine* **14** 822–827.

SHOWS, J. H., LU, W. and ZHANG, H. H. (2010). Sparse estimation and inference for censored median regression. *J. Statist. Plann. Inference* **140** 1903–1917. MR2606727

TALAGRAND, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer, Berlin. MR2133757

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809

VOLGUSHEV, S., WAGENER, J. and DETTE, H. (2014). Censored quantile regression processes under dependence and penalization. *Electron. J. Stat.* **8** 2405–2447. MR3278338

WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104** 1117–1128. MR2562007

WANG, H. J., ZHOU, J. and LI, Y. (2013). Variable selection for censored quantile regresion. *Statist. Sinica* **23** 145–167. MR3076162

WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. MR2949353

YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184. MR1325125

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448

ZHENG, Q., GALLAGHER, C. and KULASEKERA, K. B. (2013). Adaptive penalized quantile regression for high dimensional data. *J. Statist. Plann. Inference* **143** 1029–1038. MR3029230

ZHENG, Q., PENG, L. and HE, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.* **43** 2225–2258. MR3396984

ZHENG, Q., PENG, L. and HE, X. (2018). Supplement to "High dimensional censored quantile regression." DOI:10.1214/17-AOS1551SUPP.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

Q. ZHENG
DEPARTMENT OF BIOINFORMATICS AND BIOSTATISTICS
UNIVERSITY OF LOUISVILLE
LOUISVILLE, KENTUCKY 40242
USA
E-MAIL: qi.zheng@louisville.edu

L. PENG
DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATICS
EMORY UNIVERSITY
1518 CLIFTON RD, NE
ATLANTA, GEORGIA 30322
USA
E-MAIL: lpeng@emory.edu

X. HE
DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: xmhe@umich.edu