mathematics

Article

Weighted Competing Risks Quantile Regression Models and Variable Selection

Erqian Li [1], Jianxin Pan [2], Manlai Tang [3], Keming Yu [4], Wolfgang Karl Härdle [5], Xiaowen Dai [6] and Maozai Tian [7,*]

[1] College of Science, North China University of Technology, Beijing 100144, China
[2] School of Mathematics, University of Manchester, Manchester M13 9PL, UK
[3] Department of Physics, Astronomy and Mathematics, School of Physics, Engineering & Computer Science, University of Hertfordshire, Hatfield AL10 9EU, UK
[4] Department of Mathematics, College of Engineering, Design and Physical Sciences Brunel University, Uxbridge UB8 3PH, UK
[5] School of Business and Economics, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[6] School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai 201620, China
[7] Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China
* Correspondence: mztian@ruc.edu.cn

Abstract: The proportional subdistribution hazards (PSH) model is popularly used to deal with competing risks data. Censored quantile regression provides an important supplement as well as variable selection methods due to large numbers of irrelevant covariates in practice. In this paper, we study variable selection procedures based on penalized weighted quantile regression for competing risks models, which is conveniently applied by researchers. Asymptotic properties of the proposed estimators, including consistency and asymptotic normality of non-penalized estimator and consistency of variable selection, are established. Monte Carlo simulation studies are conducted, showing that the proposed methods are considerably stable and efficient. Real data about bone marrow transplant (BMT) are also analyzed to illustrate the application of the proposed procedure.

Keywords: competing risks; cumulative incidence function; bone marrow transplant; re-distribution method

MSC: 62N02

Citation: Li, E.; Pan, J.; Tang, M.; Yu, K.; Härdle, W.K.; Dai, X.; Tian, M. Weighted Competing Risks Quantile Regression Models and Variable Selection. *Mathematics* **2023**, *11*, 1295. https://doi.org/10.3390/math11061295

Academic Editors: Yiqiang Zhan and Xingrong Liu

Received: 23 January 2023
Revised: 24 February 2023
Accepted: 2 March 2023
Published: 8 March 2023

1. Introduction

In survival analysis, sometimes events fail because of a specific cause or from some other causes or competing risks. Consider the dataset of bone marrow transplant (BMT) in [1] for example, which includes 177 patients who received a stem cell transplant for acute leukemia. Whereas 56 patients in this dataset relapsed (REL), considered as the event of interest, 75 patients died from causes related to the transplant (transplant related mortality, TRM), which is considered a competing risk, as it hinders the occurrence of leukemia relapse. The other 46 patients are regarded as censored due to the end of the study. In the analysis of such a dataset, treating competing risks (TRM) as censoring cases and using usual Cox modelling may be inaccurate, as the competing risks are probably affected by covariates. To deal with such competing risks data, Ref. [2] proposed a novel semiparametric proportional hazards for the subdistribution, or PSH model, which directly analyzes the effect of covariates on the marginal probability function or cumulative incidence function (CIF). The competing risks data often occur in clinical trials containing large numbers of covariates, among which only a few have significant or essential influence on the response, generating the variable selection issues, such as the general penalized log-partial likelihood method proposed by [3].

Quantile regression introduced by [4] is widely known to more comprehensively describe the conditional distribution of response on covariates. Existing work about competing risks quantile regression includes [5], which first transforms competing risks quantile regression models to accelerated the failure model and uses an estimating equation procedure for estimation. In addition, Ref. [6] discussed the quantile regression for competing risks data with missing cause of failure. Then [7,8] developed variable selection procedures based on unbiased estimating equations with group structures and penalization methods for competing risks quantile regression models.

In the paper, in spite of the estimating equation method, we propose developing a more general method for competing risks quantile regression and expanding the weighted procedures by considering the re-distribution methods [9] for the PSH model. By transformed responses, we can rewrite the competing risks quantiles formulation as a general quantile regression objective function, then apply the constructed weights. With unbiasedness of the subgradient of this weighted objective function at the true cumulative-incidence function and coefficient proved, consistency and asymptotic normality of the penalty-free estimators are established under regularity conditions. To realize the variable selection, penalization methods such as the least absolute shrinkage and selection operator (LASSO) proposed by [10] and the adaptive LASSO (ALASSO) developed by [11] are applied to the weighted objective function, which can be easily applied with the R package. The consistency of the variable selection procedure is also established, and Monte Carlo simulation is performed to illustrate the efficiency and stability of our proposed procedures. Real data about bone marrow transplant are analyzed using our methods.

The paper is organized as follows. Our proposed weighted competing risks quantile regression model and its penalized methods are developed in Section 2, with asymptotic properties demonstrated in Section 3. Simulation studies as well as the application to the BMT data are performed in Section 4 to illustrate the performance of proposed methods.

2. Models

We take the formulation of competing risks quantile regression in [5]. In the setting of competing risks models, assume there exist K causes of failure, denoted by an observable indicator $\epsilon \in \{1, \ldots, K\}$, the same denotation as [2]. Without loss of generality, we can set $K = 2$. Let T and C denote the failure and censoring time, respectively, and we observe $X = \min(T, C)$, and censoring or risk indicator $\delta = I(T \leq C)$, where $I(\cdot)$ is an indicator function. Denote a $p \times 1$ bounded time-independent covariate vector as $\tilde{Z}$ and $Z = (1, \tilde{Z}^\top)^\top$. Assume that $\{X_i, \delta_i \epsilon_i, Z_i\}, i = 1, \ldots, n$ are independent and identically distributed observed samples.

Ref. [2] modeled the CIF for failure from cause 1 conditionally on the covariates, $F_1(t|Z) = P(T \leq t, \epsilon = 1|Z)$. They proposed the PSH model based on the formula of subdistribution hazard, which is defined as

$$\begin{aligned} \lambda_1(t|Z) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\big\{t < T \leq t + \Delta t, \epsilon = 1 \big| (T \geq t) \cup (T \leq t \cap \epsilon \neq 1), Z\big\} \\ &= \{dF_1(t|Z)/dt\}/\{1 - F_1(t|Z)\} \end{aligned}$$

in [12]. Analogue to the definition of quantile, we define the conditional quantile as $Q_k(\tau|Z) = \inf\{t : F_k(t|Z) \geq \tau\}, k = 1, \ldots, K$, where $F_k(t|Z) = P\{T \leq t, \epsilon = k\}$ is the CIF for cause k; for more details, refer to [5]. For $\tau \in [\tau_L, \tau_U]$, consider $Q_1(\tau|Z)$ to be modeled as

$$Q_1(\tau|Z) = g\{Z^\top \beta_0(\tau)\}, \tag{1}$$

where $\beta_0(\tau)$ is a $(p+1) \times 1$ coefficient vector, $g(\cdot)$ is a known monotone increasing and continuously differential bounded link function, $0 < \tau_L \leq \tau_U < 1$. With the statement in [2], if we denote $T_1^* = I(\epsilon = 1) \times T + \{1 - I(\epsilon = 1)\} \times \infty$, then T_1^* has a distribution function equal to $F_1(t|Z)$ when $t < \infty$ and a point mass $P(T^* = \infty|Z) = P(T < \infty, \epsilon \neq$

$1) = 1 - F_1(\infty|Z)$ at $t = \infty$. Then, at $\tau < F_1(\infty|Z)$, the $\tau-$quantile of T_1^* equals $F_1^{-1}(\tau|Z) = Q_1(\tau|Z) = g\{Z^\top \beta_0(\tau)\}$ under the formulation of (1).

Remark 1. *According to the formulation of T_1^*, we can see that when $\tau \geq F_1(\infty; Z)$, the $\tau-$quantile of T_1^* will become ∞, which is obvious when reviewing the definition that $F_1(t; Z) = P(T \leq t, \epsilon = 1|Z) \leq P(T \leq \infty, \epsilon = 1|Z) = F_1(\infty; Z)$ and the fact that $g(\cdot)$ is monotone increasing. This fact provides a thought about the choice of τ_U.*

With reference to [13], for proper τ, $\beta_0(\tau)$ is supposed to be the minimizer of the following expected loss function with respect to $\beta(\tau)$:

$$\beta_0(\tau) = \arg\min_{\beta(\tau)} E\rho_\tau(g^{-1}(T_1^*) - Z^\top \beta(\tau)), \tag{2}$$

where E denotes the expectation, and $\rho_\tau(u) = u\{\tau - I(u \leq 0)\}$ is called the "check" function.

In a sample scenario, we can obtain the estimator $\hat{\beta}(\tau)$ of $\beta_0(\tau)$ via minimizing the following objective function:

$$\min_{\beta(\tau)} \sum_{i=1}^{n} \rho_\tau\Big(g^{-1}(T_{1,i}^*) - Z_i^\top \beta(\tau)\Big). \tag{3}$$

2.1. Weighted Competing Risks Quantile Regression

Similar to [5], our paper first considers the case in which there are no missing data (i.e., there is no censoring). As a result, $X = T$ and $\delta = 1, \delta\epsilon = \epsilon$. As aforementioned, we can estimate $\beta_0(\tau)$ via the minimization problem (3). Because $T_{1,i}^*$ is not observed, we modify (3) to

$$\sum_{i=1}^{n} I(\epsilon_i = 1)\rho_\tau(g^{-1}(X_i) - Z_i^\top \beta(\tau)) + I(\epsilon_i \neq 1)\rho_\tau(g^{-1}(X^\infty) - Z_i^\top \beta(\tau)), \tag{4}$$

where X^∞ is any value sufficiently large to exceed all $Z_i^\top \beta(\tau)$. Then, it is not difficult to derive the negative subgradient of (4) with respect to $\beta(\tau)$.

For the censoring case, we aim to construct such a weighted quantile objective function to estimate $\beta_0(\tau)$ as follows:

$$\begin{aligned} Q(\beta(\tau), w_0) \quad = \quad & \sum_{i=1}^{n} \Big\{ w_{0i}\rho_\tau(g^{-1}(X_i) - Z_i^\top \beta(\tau)) \\ & + (1 - w_{0i})\rho_\tau(g^{-1}(X^{+\infty}) - Z_i^\top \beta(\tau)) \Big\}. \end{aligned} \tag{5}$$

The weight function is re-constructed based on competing risks analogy to [14], as follows:

$$w_{0i} = \begin{cases} 1, & \delta_i\epsilon_i = 1, \\ 0, & \delta_i\epsilon_i \neq 1, F_1(C_i|Z_i) > \tau, \\ \frac{\tau - F_1(C_i|Z_i)}{1 - F_1(C_i|Z_i)}, & \delta_i\epsilon_i \neq 1, F_1(C_i|Z_i) \leq \tau. \end{cases} \tag{6}$$

Remark 2. *In our case of competing risks quantile regression, each point contributes to the subgradient condition only via the sign of $g^{-1}(T_{1,i}^*) - Z_i^\top \beta_0(\tau)$. For data with $\delta_i\epsilon_i = 1$, we know $X_i = T_i \leq C_i, \epsilon_i = 1$, i.e., $X_i = T_{1,i}^*$, and $I(g^{-1}(T_{1,i}^*) - Z_i^\top \beta_0(\tau) < 0)$ can be observed, thus we assign a weight of 1 for this case. For data with $\delta_i\epsilon_i \neq 1$ and $F_1(C_i|Z_i) > \tau$, then $T_i > C_i, F_1(C_i|Z_i) > \tau$ or $T_i \leq C_i, \epsilon_i = 2, F_1(C_i|Z_i) > \tau$; in the first scenario, $T_{1,i}^* \geq T_i \geq X_i = C_i > g(Z_i^\top \beta_0(\tau)), I(g^{-1}(T_{1,i}^*) - Z_i^\top \beta_0(\tau) < 0) = 0$; in the second scenario, $T_i \leq C_i, \epsilon = 2, I(g^{-1}(T_{1,i}^*) - Z_i^\top \beta_0(\tau) < 0) = 0$, where we assign a weight of 0. The* *ambiguous situation is* $\delta_i\epsilon_i \neq 1$ *and* $F_1(C_i|Z_i) < \tau$, *i.e.,* $C_i \leq F_1^{-1}(\tau|Z_i) = g(Z_i^\top\beta_0(\tau))$. *If* $\delta_i = 1, \epsilon_i = 2, X_i = T_i < C_i < g(Z_i^\top\beta_0(\tau))$, *or* $I\{g^{-1}(X_i) - Z_i^\top\beta_0(\tau) < 0\} = 1$; *if* $\delta_i = 0, X_i = C_i < Z_i^\top\beta_0(\tau)$, *i.e.,* $I\{g^{-1}(X_i) - Z_i^\top\beta_0(\tau) < 0\} = 1$. *However, the* $I(T_{1,i}^* - g(Z_i^\top\beta_0(\tau)) < 0)$ *cannot be observed.*

Thus, we assign the weight $w_i(F_0) = \frac{\tau - F_1(C_i|Z_i)}{1-F_1(C_i|Z_i)}$ *for this case, where given* (Z_i, C_i),

$$\begin{aligned} &E\Big\{I(g^{-1}(T_{1,i}^*) - Z_i^\top\beta_0(\tau) < 0)|\delta_i\epsilon_i \neq 1, Z_i\Big\} \\ &= \frac{P\{\epsilon_i = 1, T_i < g(Z_i^\top\beta_0(\tau))|Z_i\} - P\{\epsilon_i = 1, T_i < C_i\}}{1 - P(T_i \leq C_i, \epsilon_i = 1|Z_i)} \\ &= \frac{\tau - F_1(C_i|Z_i)}{1 - F_1(C_i|Z_i)}. \end{aligned} \tag{7}$$

We can show that a subgradient of the weighted quantile objective function (5) with respect to $\beta(\tau)$

$$M_n(\beta(\tau), w_0) = \sum_{i=1}^{n} Z_i\Big\{\tau - w_{0i}I(g^{-1}(X_i) < Z_i^\top\beta(\tau))\Big\} \tag{8}$$

is an unbiased estimating function of $\beta_0(\tau)$.

$$\begin{aligned} &E\Big[w_{0i}I\{g^{-1}(X_i) < Z_i^\top\beta_0(\tau)\}|Z_i\Big] \\ &= E\Big(I\{\delta_i\epsilon_i = 1\}w_{0i}I\{g^{-1}(X_i) < Z_i^\top\beta_0(\tau)\}|Z_i\Big) \\ &\quad + E\Big(I\{\delta_i\epsilon_i \neq 1, F_1(C_i) > Z_i^\top\beta_0(\tau)\}w_{0i}I\{g^{-1}(X_i) < Z_i^\top\beta_0(\tau)\}|Z_i\Big) \\ &\quad + E\Big(I\{\delta_i\epsilon_i \neq 1, F_1(C_i) \leq Z_i^\top\beta_0(\tau)\}w_{0i}I\{g^{-1}(X_i) < Z_i^\top\beta_0(\tau)\}|Z_i\Big) \\ &= P\Big(\epsilon_i = 1, g^{-1}(T_i) < Z_i^\top\beta_0(\tau)|Z_i\Big) = \tau. \end{aligned}$$

Although the unbiasedness of (8) is proved with $F_1(C_i|Z_i)$ in w_{0i}, the underlying distribution $F_1(t|Z)$ or w_{0i} is unknown in practice. Here we use the IPCW [15] estimator proposed by [5] to estimate $F_1(t|Z)$,

$$\hat{F}_1(x|Z) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(X_i \leq x, \delta_i\epsilon_i = 1)}{1 - \hat{G}(X_i|Z_i)}\right\}, \tag{9}$$

where $1 - G(\cdot t|Z)$ is the survival function of C given Z, which can be estimated semiparametrically or nonparametrically. Here for simplicity, as in [2], we assume the independence of C and (T, ϵ, Z), then the Kaplan–Meier estimator in [16] could be used. Such a computation-friendly estimator (9) has been proved to behave quite well in simulation results, which should be well improved combined with more effective estimators of $F_1(t|Z)$.

By plugging (9) in the expression of w_{0i}, we can get the estimated weights $w_i(\hat{F}_1)$,

$$w_i(\hat{F}_1) = \begin{cases} 1, & \delta_i\epsilon_i = 1, \\ 0, & \delta_i\epsilon_i \neq 1, \hat{F}_1(C_i|Z_i) > \tau, \\ \frac{\tau - \hat{F}_1(C_i|Z_i)}{1 - \hat{F}_1(C_i|Z_i)}, & \delta_i\epsilon_i \neq 1, \hat{F}_1(C_i|Z_i) \leq \tau, \end{cases} \tag{10}$$

where $\hat{F}_1$ is as in (9) or replaced with other consistent estimators. Then, we obtain the weighted censoring quantile regression estimator $\hat{\beta}(\tau)$ by minimizing the weighted objective function,

$$Q(\beta(\tau),\hat{F}_1) = \sum_{i=1}^{n}\Big\{w_i(\hat{F}_1)\rho_\tau(g^{-1}(X_i)-Z_i^\top\beta(\tau)) + (1-w_i(\hat{F}_1))\rho_\tau(g^{-1}(X^{+\infty})-Z_i^\top\beta(\tau))\Big\}. \tag{11}$$

2.2. Variable Selection Procedure

To select important variables, a penalty function is added to the weighted objective function (11) to obtain the penalized estimator $\tilde{\beta}(\tau)$:

$$Q_p(\beta(\tau),w_i(\hat{F}_1)) = \sum_{i=1}^{n}\Big\{w_i(\hat{F}_1)\rho_\tau(g^{-1}(X_i)-Z_i^\top\beta(\tau)) + (1-w_i(\hat{F}_1))\rho_\tau(g^{-1}(X^{+\infty})-Z_i^\top\beta(\tau))\Big\} + \sum_{j=1}^{p}p_\lambda(|\beta_j(\tau)|), \tag{12}$$

where $p_\lambda(\cdot)$ can be LASSO, adaptive LASSO, and so on.

For LASSO and ALASSO penalty, we can easily write $p_\lambda(|\beta_j|) = \lambda_n|\hat{\beta}_j|^{-\gamma}$, where $|\hat{\beta}_j|$ is the jth element of the initial consistent unpenalized estimator. We choose $\gamma = 0$ for LASSO and $\gamma = 1$ for ALASSO. The minimization of (12) and (11) can be directly solved with the R package quantreg without linear programming, leading our proposed methods to conveniently applicable tools.

3. Theoretical Property

To establish the asymptotic results in this paper, we require the following assumptions:

A1 The covariate Z is bounded in probability. There exists a constant K_z such that $E\|Z\|^3 \le K_z$, and $E(ZZ^\top)$ is a positive definite $(p+1)\times(p+1)$ matrix.

A2 The functions $F_1(t|Z)$ and $G(t)$ have first derivatives with respect to t, denoted as $f_1(t|Z)$ and $g_0(t)$, which are uniformly bounded away from infinity. Additionally, $F_1(t|Z)$ and $G(t)$ have bounded (uniformly in t) second-order partial derivatives with respect to Z.

A3 For β in the neighborhood of $\beta_0(\tau)$, $E(ZZ^\top g'(Z^\top\beta)f_1(g(Z^\top\beta)|Z)\{1-G(g(Z^\top\beta))\})$ and $E(ZZ^\top g'(Z^\top\beta)g_0(g(Z^\top\beta)))$ are positive definite.

Assumption A1 states some tail and moment conditions on the covariate Z, which are standard for the quantile regression. Assumption A2 is needed for the local Kaplan–Meier estimator. It allows us to obtain the local expansions of $F_1(t|z)$ and $G(t)$ in the neighborhood of $Z^\top\beta_0(\tau)$ in order to obtain the uniform consistency and the linear representation of $\hat{F}_1(t|Z)$. Assumption A3 ensures that the expectation of the estimating function $E\{M_n(\beta,F_1)\}$ has a unique zero at $\beta_0(\tau)$, and it is needed to establish the asymptotic distribution of $\hat{\beta}(\tau)$.

C1 There exists $\nu > 0$ such that $P(C=\nu) > 0$ and $P(C>\nu)$=0.

C2 $\beta_0(\tau)$ is Lipschitz continuous for $\tau \in [\tau_L,\tau_U]$.

C3 $P(\epsilon = 1|Z) < 1$ a.s.

Assumptions C1 and C2 are regularity conditions for competing risks quantile regressions. Assumption C3 is easily satisfied for the situation of competing risks; otherwise, it will turn out to be a standard Cox model.

Theorem 1. *Assume that triples $\{Z_i, X_i, \delta_i\epsilon_i\}, i = 1,\cdots,n$ constitute an i.i.d. multivariate random sample and that the censoring variable C_i is independent of T_i conditionally on the covariate Z_i. Under model (1) and assumptions A1–A3, C1–C3,*

$$\hat{\beta}(\tau) \to \beta_0(\tau) \tag{13}$$

in probability as $n \to \infty$.

Theorem 2. *Under the assumptions of Theorem 1 and $r < 1/4$, we have*

$$n^{1/2}(\hat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{D} \mathrm{N}(0, \mathbf{\Gamma^{-1} V \Gamma^{-1}}), \tag{14}$$

where

$$\Gamma^{-1} = E[ZZ^\top g'(Z^\top \beta_0(\tau))\{1 - G(g(Z^\top \beta_0(\tau)))\} f_1(g(Z^\top \beta_0(\tau))|Z)], \tag{15}$$

and

$$\mathbf{V} = \mathrm{Cov}(m_i(\beta_0, F_1) + (1-\tau)\phi_i), \tag{16}$$

with $m_i(\beta_0, F_1) = Z_i\{\tau - w_i(F_1) I(X_i < g(Z_i^\top \beta_0(\tau)))\}$, ϕ_i defined in Equation (A9).

Theorems 1 and 2 established the consistency and asymptotic normality of the unpenalized estimator $\hat{\beta}(\tau)$. We then establish the property of consistency in variable selection of the proposed penalized estimator $\tilde{\beta}(\tau)$. Let $\mathcal{A}(\tau) = \{j : \beta_{0j} \neq 0\}$ and $\mathcal{A}^c(\tau) = \{j : \beta_{0j}(\tau) = 0\}$.

Theorem 3. *If A1–A3, C1–C3 hold, and if $n^{-1/2}\lambda_n \to 0$ and $n^{(\gamma-1)/2}\lambda_n \to \infty$, then*

$$P\big(\{j : \tilde{\beta}_j(\tau) \neq 0\} = \mathcal{A}(\tau)\big) \to 1 \quad \text{as} \quad n \to \infty.$$

Theorem 3 states that the proposed procedure is able to select the correct model with probability approaching one. By the remark of Theorem 2 of [14], the oracle properties are satisfied by the proposed estimators.

The proofs are presented in Appendix A.

4. Numerical Studies

4.1. Monte Carlo Simulation

We conduct Monte Carlo simulations to evaluate the performance of the proposed methods and consider the data-generating ways as in [5] with a larger dimension of covariates.

We generate (T, ϵ) satisfying $P(\epsilon = 1|Z) = p_0 I(Z_2 = 0) + p_1 I(Z_2 = 1)$, $P(T \leq t|\epsilon = 1, Z) = \Phi(\log t - \gamma_0^\top Z)$, and $P(T \leq t|\epsilon = 2, Z) = \Phi(\log t - \alpha_0^\top Z)$, where $\Phi(\cdot)$ denotes the standard normal distribution function, $p_0 = 0.8$, $p_1 = 0.6$, γ_0 and α_0 are true parameters in the model above. Set $\gamma_0 = (-2, -2.5, 2, -2.4, 0, \cdots, 0)$ while $\alpha_0 = -\gamma_0$. Then

$$\begin{aligned} \log Q_1(\tau|Z) &= \Phi^{-1}\left(\frac{\tau}{p_0}\right) + \gamma_0^{(1)} Z_1 + \left\{\gamma_0^{(2)} + \Phi^{-1}\left(\frac{\tau}{p_1}\right) - \Phi^{-1}\left(\frac{\tau}{p_0}\right)\right\} Z_2 \\ &\quad + \gamma_0^{(3)} Z_3 + \gamma_0^{(4)} Z_4 \end{aligned}$$

where Z_j is the jth component of covariate Z, and $\gamma_0^{(j)}$ is the jth component of γ_0. Then the estimated coefficient in model (1) is

$$\beta_0(\tau) = (\Phi^{-1}\left(\frac{\tau}{p_0}\right), \gamma_0^{(1)}, \gamma_0^{(2)} + \Phi^{-1}\left(\frac{\tau}{p_1}\right) - \Phi^{-1}\left(\frac{\tau}{p_0}\right), \gamma_0^{(3)}, \gamma_0^{(4)}, 0, \cdots, 0).$$

Thus the true number of non-zero coefficients is 5 for $\tau \neq 0.4$ and 4 for $\tau = 0.4$ due to $\Phi^{-1}\left(\frac{0.4}{p_0}\right) = 0$.

In simulations, we set the number of irrelevant predictors to be $s = \#\{j : \beta_{0j} \neq 0\} = 30$, the sample size to be $n = 200$. For the structure of the covariance matrix for covariates, we consider $\Sigma_{1,ij} = \rho$,where $\rho = 0, 0.25, 0.5, 0.75$.

We generate the covariate vector $Z = (Z_1, Z_2, Z_3, \cdots, Z_p)^\top$ as follows: $Z_1 \sim$ Unif(0,1) and $Z_2 \sim$ Bernoulli(0.5), $Z_j \sim N(0, \Sigma)$, $j = 3, \cdots, p$. For each scenario, the simulation is repeated 500 times. The censoring rate average is 36%.

We use the following criteria to evaluate the performances: the ratio of number of relevant variables correctly selected to true number of relevant variables (TPr) defined as TPr $= \frac{\#\{\{j:|\hat{\beta}_j(\tau)|\neq 0\}\cap\{j:|\beta_{0j}(\tau)|\neq 0\}\}}{\#\{j:|\beta_{0j}(\tau)|\neq 0\}}$, the ratio of number of irrelevant variables incorrectly selected to true number of irrelevant variables (FPr) defined as FPr $= \frac{\#\{\{j:\hat{\beta}_j(\tau)\neq 0\}\cap\{j:|\beta_{0j}|=0\}\}}{\#\{j:|\beta_{0j}|=0\}}$, the absolute error $P_1 = \sum_{j=1}^{p} |\hat{\beta}_j(\tau) - \beta_{0j}(\tau)|$, and the squared error $P_2 = \sum_{j=1}^{p} |\hat{\beta}_j(\tau) - \beta_{0j}(\tau))|^2$. The closer TPr is to 1 and FPr is to 0, the better. Both TPr and FPr range from 0 to 1, thus we present them together in Figure 1 for comparison.

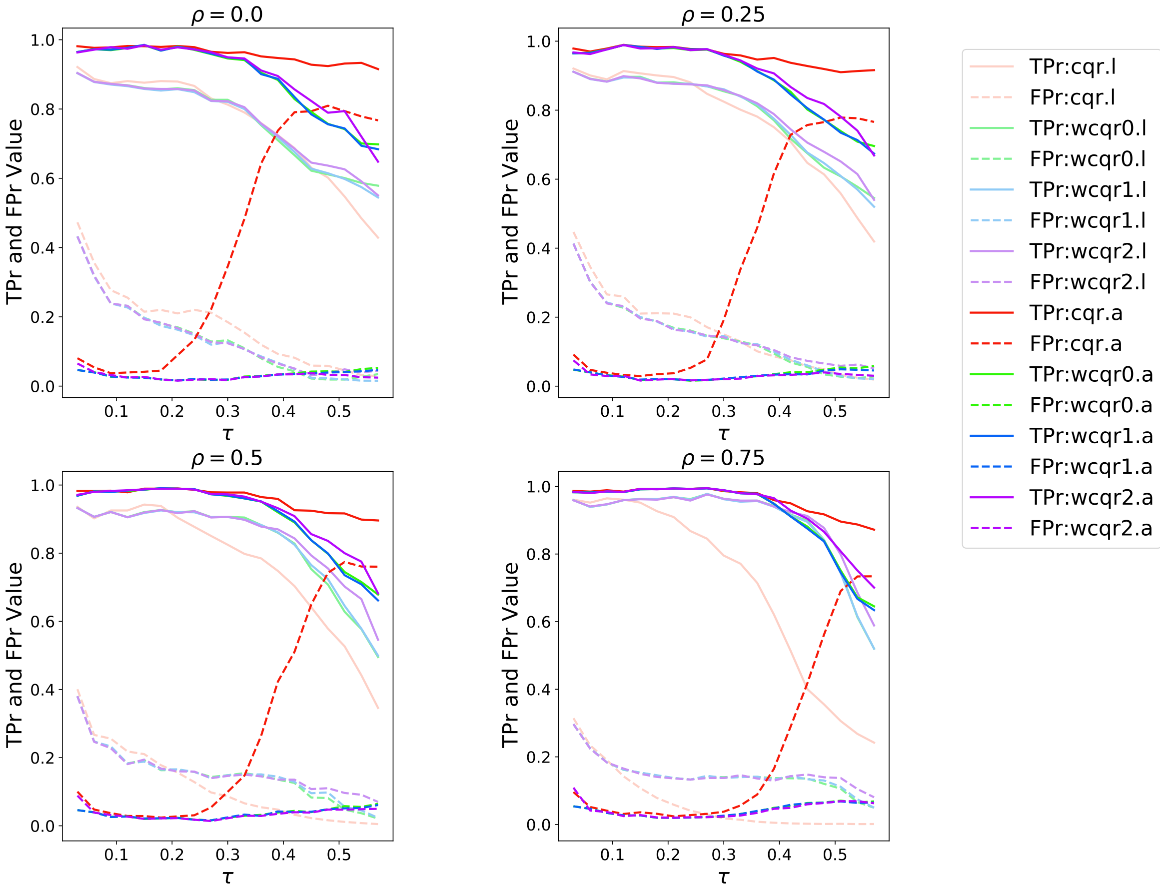

Figure 1. Case $n = 200, s = 30, p_0 = 0.8, p_1 = 0.6$. Comparison of TPr and FPr for four levels of ρ. In each subplot, the Y axis reports the TPr and FPr values at different τ. The solid line is TPr and the dashed line is FPr. Four colors are used to represent the methods: red for wcqr, green for wcqr0, blue for wcqr1, and purple for wcqr2. Light colors represent the LASSO penalty and dark colors are the ALASSO penalty for all methods.

We compare our proposed weighted estimators with the estimated estimator of competing risks quantile regression model proposed in [8], denoted as wcqr and cqr, respectively, implying the weighting method or not. In simulation tables, we use cqr.l and cqr.a to represent cqr estimators with LASSO and ALASSO penalty, respectively. Similarly, our estimators, denoted as wcqr$i$.l and wcqr$i$.a, $i = 0, 1$ stands for administrative censoring where C is known and randomly right censoring cases where X is in place of C, respectively; wcqr2 uses a different weight:

$$w_i(F_1) \quad = \quad \begin{cases} 1 & \delta_i \epsilon_i = 1 \\ \frac{\tau - \hat{F}_1(C_i)}{1 - \hat{F}_1(C_i) - \hat{F}_2(C_i)} & \delta_i = 0, \hat{F}_1(C_i) < \tau \\ 0 & \text{otherwise.} \end{cases}$$

As the weight above involves the estimation of $F_2(t|Z) = P(T \leq t, \epsilon = 2|Z)$, which probably is complicated in practical circumstances, we only use it for comparison in simulations. Here in wcqr2, we apply a similar estimating method of F_1 to F_2.

Alhough our theoretical results are not based on these two estimators of w_i, most simulation results show that wcqr0 and wcqr1 are considerably close, as the weight is only different at $\delta = 1$, suggesting good estimates in large censoring rates. Research about massive competing risks data with enormous censored observations will appear in our future work.

Before the variable selection, we also conduct the simulation for unpenalized estimators. In this case, we use $\gamma_0 = (1, -1.5, -0.5)$, $p_0 = 0.8$, $p_1 = 0.6$, and $\rho = 0$. We repeat this 1000 times and compare the empirical bias (EmpBias) and average coverage probabilities based on 95% confidence intervals computed with empirical variance. The results are summarized in Table 1, where in lower quantiles, the cqr method shows extreme excellence, whereas in high quantiles, it displays some instability. For weighted methods, though inferior to cqr in lower quantiles, these methods still behave well in most simulations, especially wcqr1 and wcqr2. The average coverage probabilities display similar patterns; cqr behaves well until $\tau < 0.4$. In relatively high quantiles such as $\tau = 0.5$, wcqr2 behaves the best for most coefficients.

Table 1. Bias and empirical coverage; $n = 300, \rho = 0, p_0 = 0.8, p_1 = 0.6$.

τ	Method	Bias				EmpCoverage			
		β_1	β_2	β_3	β_4	β_1	β_2	β_3	β_4
0.1	cqr	−0.012	−0.003	0.003	0.003	0.952	0.953	0.953	0.958
	wcqr1	−0.027	−0.026	0.031	0.014	0.950	0.952	0.950	0.949
	wcqr2	−0.027	−0.032	0.036	0.018	0.948	0.952	0.946	0.950
0.2	cqr	0.000	−0.015	−0.007	0.008	0.946	0.950	0.943	0.952
	wcqr1	−0.024	−0.061	0.042	0.032	0.948	0.944	0.949	0.943
	wcqr2	−0.019	−0.081	0.055	0.040	0.949	0.940	0.952	0.940
0.3	cqr	−0.010	−0.023	0.008	0.014	0.947	0.949	0.956	0.953
	wcqr1	−0.042	−0.108	0.095	0.055	0.938	0.952	0.944	0.943
	wcqr2	−0.027	−0.131	0.104	0.065	0.939	0.949	0.934	0.941
0.4	cqr	0.219	−1.244	−0.069	0.316	0.999	0.999	0.999	0.999
	wcqr1	−0.065	−0.203	0.168	0.121	0.947	0.943	0.931	0.937
	wcqr2	0.007	−0.153	0.111	0.095	0.949	0.939	0.941	0.922
0.5	cqr	−5.878	−25.526	−12.526	10.139	0.975	0.957	0.967	0.953
	wcqr1	−0.313	−0.632	0.594	0.280	0.965	0.943	0.963	0.914
	wcqr2	0.144	0.034	0.041	−0.007	0.913	0.955	0.952	0.953

With moderate dimensions of covariates ($s = 30$), Figure 1 presents the TPr and FPr values evaluated for $n = 200$ and $\tau \in (0, 0.6)$ and four ρs for Σ_1, respectively. Generally speaking, we can observe that almost all selection performances appear to decline as τ increases, and with higher TPr and lower FPr, ALASSO penalized methods are overall superior to LASSO methods. Specifically, in quantiles lower than 0.4, with ALASSO penalty, cqr and wcqr both have good performances for identification of important variables, with TPr close to 1. Compared to the ALASSO method, LASSO methods have higher FPr values and tend to select much more irrelevant variables. In Figure 1, wcqr estimators display comparable performance with cqr estimators according to high TPr and low FPr at moderate $\tau \in (0, 0.35)$. At higher quantiles, although a little bit inferior to cqr estimators in TPr, wcqr estimators have very low FPr values despite a rapid increase of FPr for cqr estimators, which means wcqr estimators have a strong ability to drop irrelevant variables as well as select correct variables even when cqr estimators almost fail in particularly high quantiles. We should state that in all simulations, wcqr estimators present quite stable performances in higher quantiles and higher dimensions. The decline of performance with increasing τ can be explained by a higher τ that is approaching the probability $P(\epsilon = 1|Z)$, which induces larger biases. In addition, it is notable that TPr has a very small decrease when ρ increases except for cqr.l, which has a large decrease, since when the correlation of covariates increases, it is more difficult for identification. Even when $\rho = 0.75$, the wcqr estimators with ALASSO behave quite well in simulations.

Figure 2 shows the P1 and P2 performances for the eight methods, and the two values for cqr estimators are too large to be displayed in the plot. In contrast, wcqr estimators stably indicate a decrease from 0.1 to about 0.27 and an increase from 0.3 to 0.6. It can be explained that in low quantiles, few ambiguous cases are used for estimation, which causes insufficient use of information; whereas in high quantiles, where more ambiguous observations are weighted, the accuracy of weights will affect the estimation performance. The improvement of estimation for w_{0i} can be an investigation in the future.

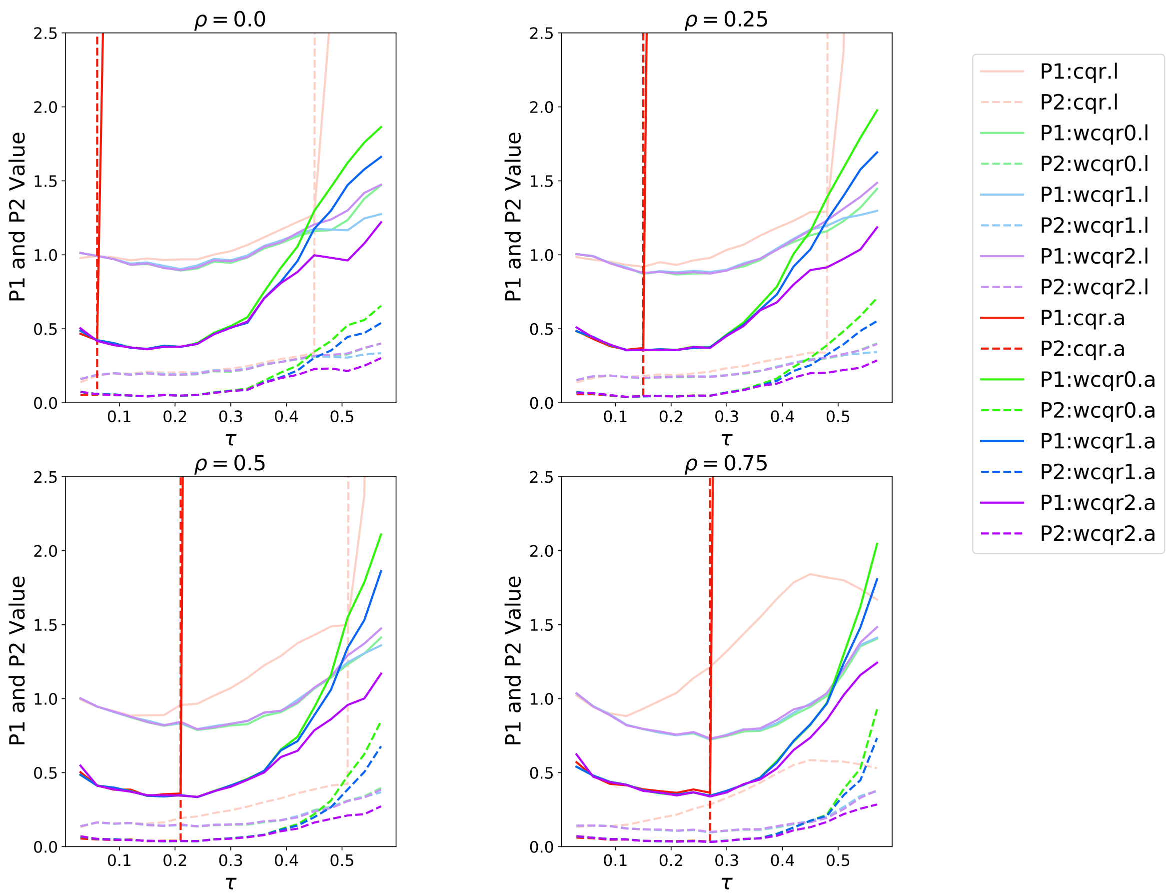

Figure 2. Case $n = 200, s = 30, p_0 = 0.8, p_1 = 0.6$. Comparison of P1 and P2 for four levels of ρ. In each subplot, the Y axis reports the P1 and P2 values at different τ. The solid line is P1 and the dashed line is P2. Four colors are used to represent the methods: red for cqr, green for wcqr0, blue for wcqr1, and purple for wcqr2. Light colors represent the LASSO penalty and dark colors are the ALASSO penalty for all methods.

We also present other simulation results in Figures S1–S11 in the supplementary material. Figures S1 and S2 show the TPr, FPr, P1, and P2 for $s = 20$ and 50, respectively. We can observe that, in Figure S1, the TPr of wcqr estimators with ALASSO are above 0.9 except for very high quantiles, indicating stability with low FPr compared to cqr estimators; in Figure S2, the tendency remains but the selection performance is inferior, although TPr still stays higher than 0.8 at $\tau = 0.4$. We also conduct the case when $s = 100, n = 100$, which means the number of predictors exceeds the sample size. In this case, cqr estimators fail due to singular design matrix as well as the ALASSO estimator. We discover, surprisingly, that our wcqr estimators still work and behave quite well, as illustrated in Figure S3. Numerical studies for $s = 10$ and $\gamma_0 = (-2, -2.5, 0.5, 0, \cdots, 0)$ are also discussed in the supplementary material, illustrated by Figures S4–S11. For the structure of the covariance matrix, we consider another kind of setup: $\Sigma_{2,ij} = \rho^{|i-j|}$. We also consider a different choice for p_0 and p_1 as 0.6 and 0.45, respectively, in order to test the performance under a different probability of $P(\epsilon_i = 1)$. In addition, we also simulate the heavy-tailed distributions $t(3)$ instead of Gaussian distribution for $P(T \leq t|\epsilon = 1, Z)$. Figures S4–S7 show that the ALASSO penalty significantly decreases the FP for both estimators, which suggests the superiority of ALASSO. Our estimators behave fairly close to the cqr estimator in most cases. Although the TP of our estimators behave slightly worse, the FP shows a relatively better performance. Not only does wcqr shows smaller deviation about estimated coefficients, but it also shows great stability, especially for the ALASSO penalty, in the case of higher quantiles $\tau = 0.5$. This shows the meaningful application of our estimators in high quantiles. Figure S8 represents the case of $n = 400$, where the performances of all criteria are greatly improved.

Figure S9 shows the performance for Σ_2, which presents slightly better results than the case of Σ_1. Figure S10 is for a different pair of $(p_0, p_1) = (0.6, 0.45)$, and our τs ranges from 0 to 0.4, and $\tau = 0.3$ turns out to be the quantile of 3 nonzero coefficients, which fits our simulation results. Figure S11 simulates t(3) distribution in place of standard normal distribution, displaying that our estimators behave significantly well for heavy-tailed distributions.

To conclude, wcqr estimators behaves comparably with the cqr estimator, with slightly worse performance for TP but better for FP. Interestingly, for the higher correlations and higher quantiles and heavy-tailed distribution, the superior performance the wcqr estimators display show good potential applicability to more complex data and higher quantiles.

4.2. Real Data Analysis

In this subsection, we use the BMT dataset in [1] for practical application. As the simulation illustrates, wcqr estimators display more stability to the complexity of data and high quantiles than existing cqr estimators, which motivates us to conduct the data analysis with our methods.

In this dataset, a total of 177 patients received a stem cell transplant for acute leukemia. The failure event is relapse (REL, 56 patients), and death from causes related to the transplant (transplant related mortality, TRM, 75 patients) is the competing risk. Forty-six patients are censored, thus the censoring rate is 26%. Covariates that affect REL and TRM includes sex, disease (lymphoblastic or myeloblastic leukemia), phase at transplant (Relapse, CR1, CR2, CR3), source of stem cells (bone marrow and peripheral blood, coded as BM+PB, or peripheral blood, coded as PB), and age. The link function is assumed to be exponential.

Figures 3–5 report the numbers of selected variables as well as coefficient estimates by our weighted estimators compared with penalized quantile estimating equations proposed by [8] and the penalty-free methods with τ ranging from 0 to 0.4.

From the figure we can see mainly our estimators select similar numbers of variables to cqr estimators at lower quantiles, but in higher quantiles, the wcqr estimators lie between the cqr-LASSO and cqr-ALASSO. For the intercept, in lower quantiles, five estimators appears conincident with one another, although cqr-ALASSO estimators tend to be unstable, whereas wcqr estimators shows stability here. For age, all estimators regard this variable as unimportant, except that the two LASSO estimators probably overestimate the importance. For sex:F, almost all estimators shrink the corresponding coefficients to zero. The ALASSO estimators tend to treat D:AML as an unimportant variable, except for quantiles around 0.1. For phase:CR1 and phase:CR2, all estimators tend to select them in lower quantiles, but wcqr tends to select phase:CR1 at higher quantiles larger than 0.21 but neglects phase:CR2 from 0.22 to 0.27. For phase:CR3, all the estimators show analogue performances but with slight shifts. For source:PB, the wcqr estimators perform more stably than cqr for all quantiles. The estimations for F_1 based on the five methods are placed in Figure 6.

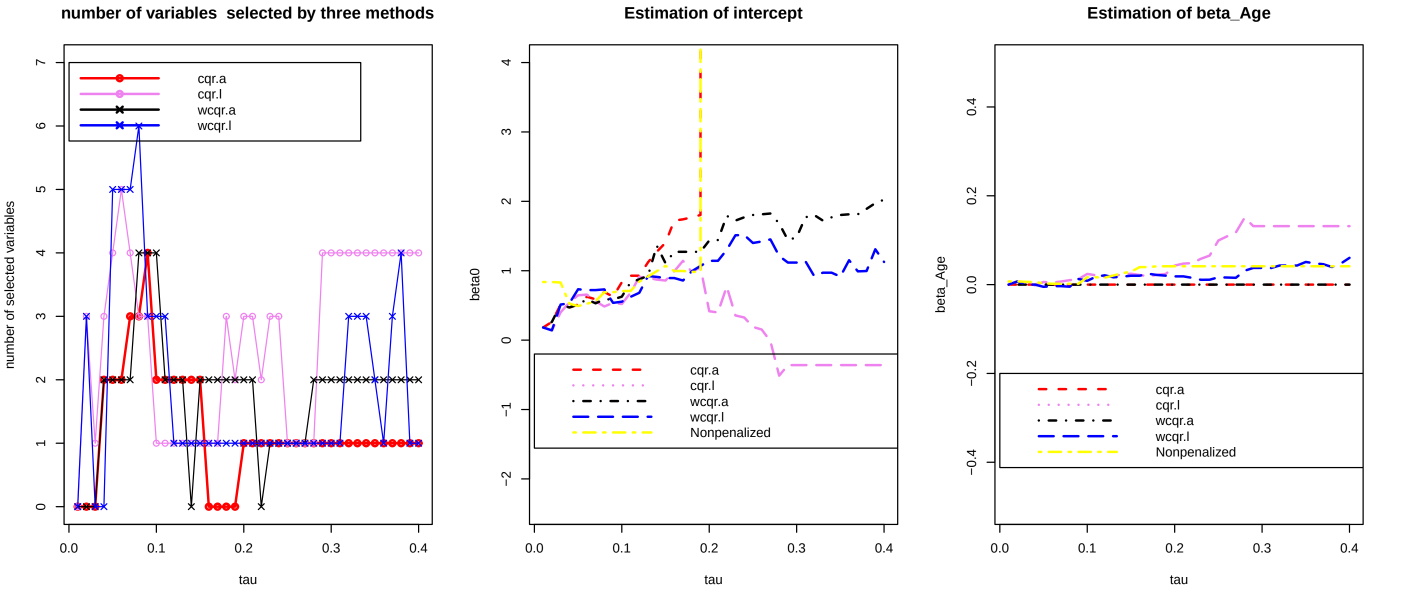

Figure 3. Variable selection and estimation results for intercept and β_{Age}. The Y axis reports the coefficient values at different τ. Various colors of lines represent eight methods.

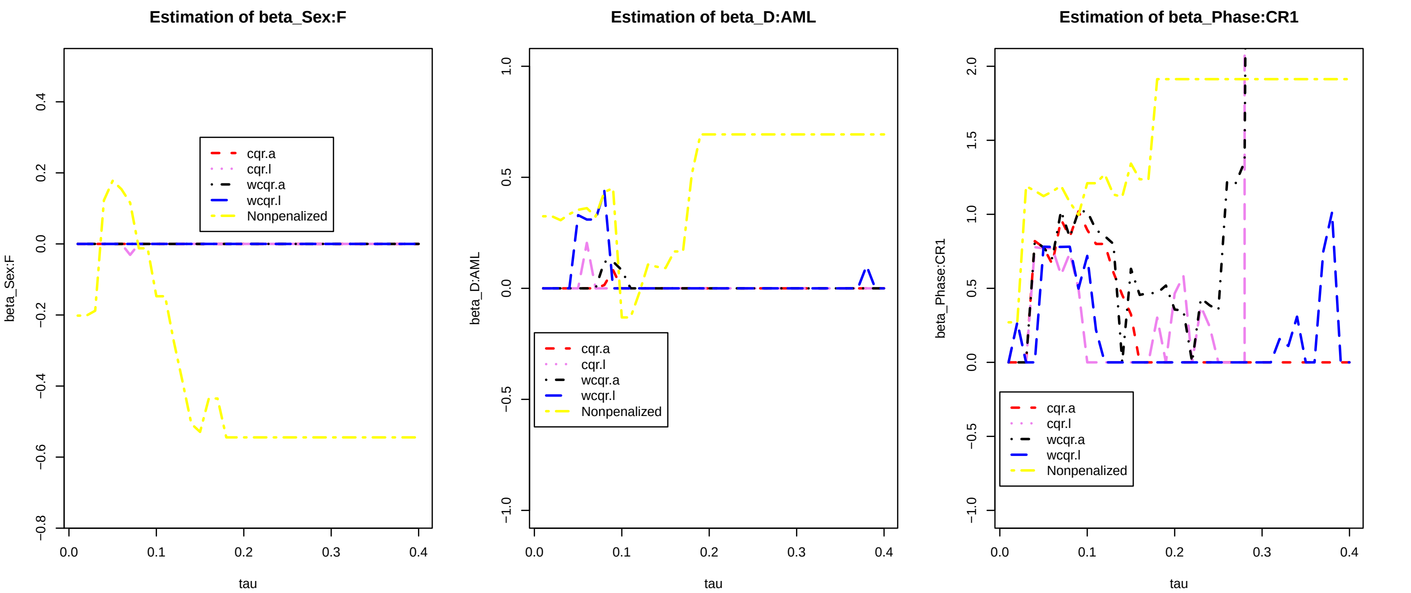

Figure 4. Estimation results for $\beta_{\text{Sex:F}}$, $\beta_{\text{D:AML}}$, and $\beta_{\text{Phase:CR1}}$.The Y axis reports the coefficient values at different τ. Various colors of lines represent eight methods.

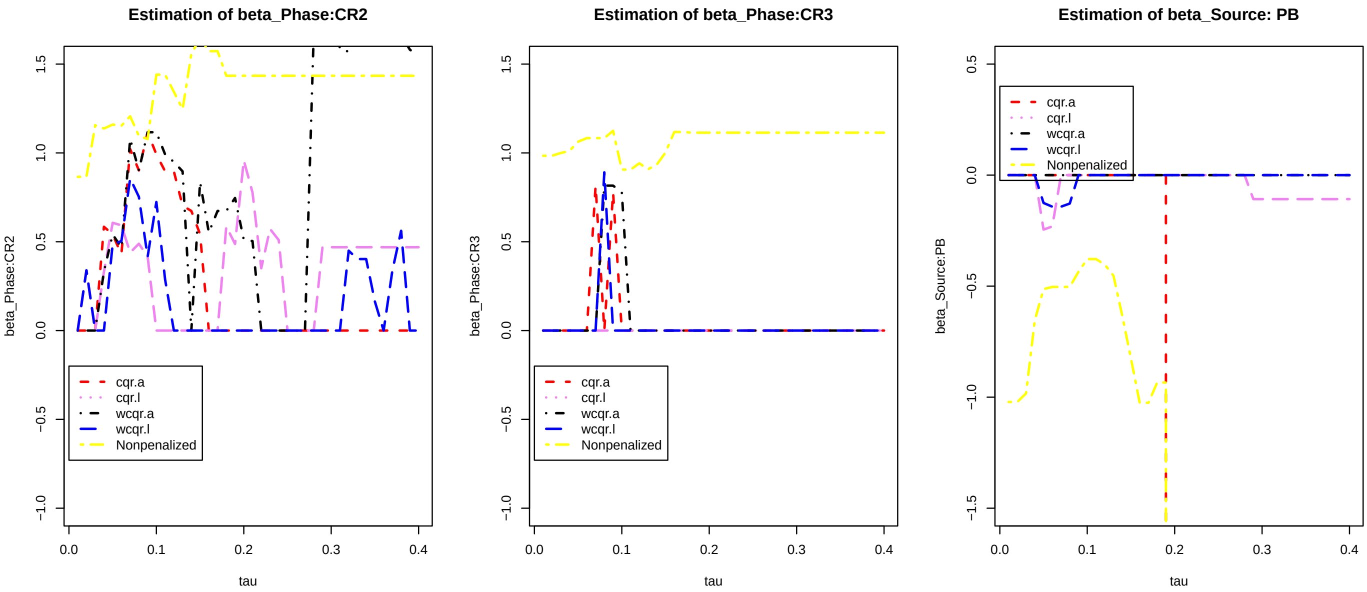

Figure 5. Estimation results for $\beta_{\text{Phase:CR2}}$, $\beta_{\text{Phase:CR3}}$ and $\beta_{\text{Source:PB}}$ .The Y axis reports the coefficient values at different τ. Various colors of lines represent eight methods.

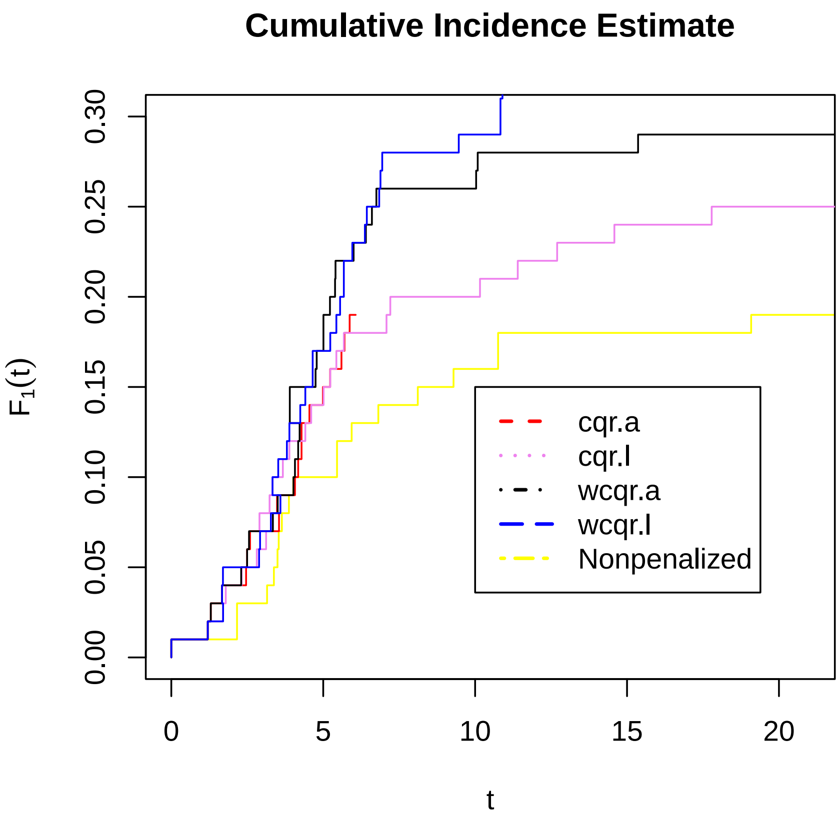

Figure 6. Estimation of F_1. The Y axis reports the estimated values of $F_1(t)$ at different t. Various colors of lines represent eight methods.

To conclude, our wcqr estimators present stability and keep similar performances to the results of cqr estimators. More importantly, our weighted estimates provide a relatively general objective function for researchers to directly use R packages for application.

5. Conclusions

In this paper, we proposed a weighted method for competing risks quantile regression model to transform the estimating equation to a common weighted objective function and applied the LASSO and ALASSO penalization for variable selection. We established the consistency and asymptotic normality for penalty-free estimators as well as the consistency of variable selection. Monte Carlo simulations were conducted for several scenarios, presenting good variable selection performance and stability. Finally, a real dataset was utilized to illustrate the application of our methods, which is comparable with other methods.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math11061295/s1, Figure S1: Case $n = 200, s = 20, p_0 = 0.8, p_1 = 0.6$. Reports of TPr, FPr, P_1 and P_2 at different τ. Four colors are used to represent the methods: red for cqr, green for wcqr0, blue for wcqr1 and purple for wcqr2. Light colors represent LASSO penalty and dark colors for ALASSO penalty. Figure S2: Case $n = 200, s = 50, p_0 = 0.8, p_1 = 0.6$. Reports of TPr, FPr, P_1 and P_2 at different τ. Four colors are used to represent the methods: red for cqr, green for wcqr0, blue for wcqr1 and purple for wcqr2. Light colors represent LASSO penalty and dark colors for ALASSO penalty. Figure S3: Case $n = 100, s = 100, p_0 = 0.8, p_1 = 0.6$, Reports of TPr, FPr, P_1 and P_2 at different τ. Various colors of line represent eight methods respectively. Figure S4: Case $n = 200, p_0 = 0.8, p_1 = 0.6$. Comparision of TP and FP for four levels of ρ. In each subplot, the Y axis reports the TP values at different τ. Various colors of line represent eight methods respectively. Figure S5: Case $n = 200, p_0 = 0.8, p_1 = 0.6$, Plots of FP for four levels of ρ. In each subplot, the Y axis reports the FP values at different τ. Various colors of line represent eight methods respectively. Figure S6: Plots of P_1 for four levels of ρ. In each subplot, the Y axis reports the P_1 values at different τ. Various colors of line represent eight methods respectively. $n = 200, p_0 = 0.8, p_1 = 0.6$. Figure S7: Plots of P_2 for four levels of ρ. In each subplot, the Y axis reports the P_2 values at different τ. Various colors of line represent eight methods respectively, $n = 200, p_0 = 0.8, p_1 = 0.6$. Figure S8: Case $n = 400, \rho = 0.5, p_0 = 0.8, p_1 = 0.6$. Reports of TP, FP, P_1 and P_2 at different τ. Various colors of line represent eight methods respectively. Figure S9: Case $n = 400, \rho = 0.5, p_0 = 0.8, p_1 = 0.6, \Sigma_2$. Reports of TP, FP, P_1 and P_2 at different τ. Various colors of line represent eight methods respectively. Figure S10: Case $n = 400, \rho = 0.5, p_0 = 0.6, p_1 = 0.45, \Sigma_1$. Reports of TP, FP, P_1 and P_2 at different τ. Various colors of line represent eight methods respectively. Figure S11: Case $n = 400, \rho = 0.5, p_0 = 0.8, p_1 = 0.6, \Sigma_2, t(3)$. Reports of TP, FP, P_1 and P_2 at different τ. Various colors of line represent eight methods respectively.

Author Contributions: Conceptualization, E.L., J.P., M.T. (Manlai Tang), K.Y., W.K.H., X.D. and M.T. (Maozai Tian); Methodology, E.L.; Software, E.L.; Validation, E.L.; Formal analysis, E.L.; Investigation, E.L.; Resources, E.L.; Data curation, E.L., J.P., M.T. (Manlai Tang), K.Y., W.K.H., X.D. and M.T. (Maozai Tian); Writing—original draft, E.L.; Writing—review & editing, E.L., J.P., M.T. (Manlai Tang), K.Y., W.K.H., X.D. and M.T. (Maozai Tian); Visualization, E.L.; Supervision, J.P., M.T. (Manlai Tang), K.Y., W.K.H. and M.T. (Maozai Tian); Project administration, E.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Funds of China (Grant No. 12101015), Scientific Research Foundation of North China University of Technology (No. 110051360002), the Fundamental Research Funds for Beijing Universities, NCUT (NO.110052971921/007), National Natural Science Foundation of China (No. 11861042), and the China Statistical Research Project (No. 2020LZ25).

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: 10.1038/bmt.2009.359.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Technical Details of Proofs

To simplify the presentation, we omit τ in such expressions as $\beta(\tau)$.

Since the weights w_i depend on F_1^* , we take w_i as $w_i(F_1^*)$. Additionally, we define $M_n(\beta, F_1^*) = n^{-1}\sum_{i=1}^n m_i(\beta, F_1^*)$ as the subgradient of the weighted quantile objective function (11), where

$$\begin{aligned} m_i(\beta, F_1^*) &= Z_i\{\tau - w_i(F_1^*)I(g^{-1}(X_i) \le Z_i^\top\beta)\} \\ &= Z_i\Big(\tau - I\{\epsilon_i = 1, T_i \le C_i, C_i \le g(Z_i^\top\beta)\} \\ &\quad - I\{\epsilon_i = 1, T_i \le C_i, g^{-1}(T_i) \le Z_i^\top\beta, C_i > g(Z_i^\top\beta)\} \\ &\quad - \frac{\tau - F_1^*(C_i)}{1 - F_1^*(C_i)}\Big[I\{F_1^*(C_i) \le \tau, C_i \le g(Z_i^\top\beta)\}(1 - I\{T_i \le C_i, \epsilon_i = 1\})\Big] \\ &\quad - \frac{\tau - F_1^*(C_i)}{1 - F_1^*(C_i)}I\{\epsilon_i = 2, F_1^*(C_i|Z_i) \le \tau, T_i \le g(Z_i^\top\beta), C_i > g(Z_i^\top\beta)\}\Big) \end{aligned}$$

Let $M(\beta, F_1^*) = E\{m_n(\beta, F_1^*)\} = E\{Z(\tau - H(g(Z^\top\beta)) - R(\beta, F_1^*) - J(\beta, F_1^*))\}$, where

$$\begin{aligned} H(t|Z) &= \int_{-\infty}^t F_1(u)g_0(u)\mathrm{d}u + (1 - G(t))F_1(t|Z), \\ R(\beta, F_1^*) &= E_{C|Z}\frac{\tau - F_1^*(C)}{1 - F_1^*(C)}I\{F_1^*(C) \le \tau, C \le g(Z^\top\beta)\}(1 - I\{T \le C, \epsilon = 1\}) \\ &= \int_0^{g(Z^\top\beta)} g_0(u)I\{F_1^*(u) \le \tau\}(1 - F_1(u|Z))\frac{\tau - F_1^*(u)}{1 - F_1^*(u)}\mathrm{d}u, \\ J(\beta, F_1^*) &= E_{C|Z}I\{\epsilon = 2, F_1^*(C) \le \tau, T \le g(Z^\top\beta), C > g(Z^\top\beta)\}\frac{\tau - F_1^*(C)}{1 - F_1^*(C)} \\ &= (F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z))\int_{g(Z^\top\beta)}^\infty I\{F_1^*(u) \le \tau\}\frac{\tau - F_1^*(u)}{1 - F_1^*(u)}g_0(u)\mathrm{d}u \end{aligned}$$

where $g_0(u)$ is the density of censoring variable C conditionally on Z, and $F_0(t|Z) = P(T \le t|Z)$. It is noteworthy that $J(\beta_0, F_1) \equiv 0$, and it is easy to derive that $\mathrm{M}(\beta_0, F_1) \equiv 0$.

Lemma A1. *Assume assumptions A1–A3, C1–C3 hold. Then*

$$\|\hat{F}_1 - F_1\|_{\mathcal{H}} \doteq \sup_t \sup_z |\hat{F}_1(t|z) - F_1(t|z)| = o_p(n^{-1/2+r}) \tag{A1}$$

for every $r > 0$.

Remark A1. *Lemma A1 directly guarantees the consistency of our weight estimation* $w_i(\hat{F}_1)$ *to* $w_i(F_1)$*, which is the* w_{0i} *in Equation (6).*

Proof. By condition C1 and A1 and [17], Ref. [5] has developed that for every $r > 0$, $\sup_{t<\nu}|\hat{G}(t) - G(t)| = o(n^{-1/2+r})$, a.s. This, coupled with C2, implies that

$$\begin{aligned} \sup_x \Bigg\| & n^{-1}\sum_{i=1}^n \left(\frac{I\{X_i \le x\}I\{\delta_i\epsilon_i = 1\}}{1 - \hat{G}(X_i)}\right) \\ & - n^{-1}\sum_{i=1}^n \left(\frac{I\{X_i \le x\}I(\delta_i\epsilon_i = 1)}{1 - G(X_i)}\right)\Bigg\| = o(n^{-1/2+r}), a.s. \end{aligned} \tag{A2}$$

Simultaneously , for $t < \nu$, $1 - G(t)$ is uniformly bounded away from 0, thus by Chebyshev's inequality, for every $r > 0$,

$$P\left\{ n^{1/2-r} \left| n^{-1} \sum_{i=1}^{n} \left(\frac{I\{X_i \leq x\} I(\delta_i \epsilon_i = 1)}{1 - G(X_i)} \right) - n^{-1} \sum_{i=1}^{n} E\left(\frac{I\{X_i \leq x\} I(\delta_i \epsilon_i = 1)}{1 - G(X_i)} | Z_i \right) \right| \geq \varepsilon \right\} \leq \frac{n^{-2r} \mathrm{Var}(I\{X_i \leq x\} I(\delta_i \epsilon_i = 1) | Z_i)}{\varepsilon^2} \to 0, n \to \infty,$$

which holds for any x, that is

$$\sup_{x,z} \left\| n^{-1} \sum_{i=1}^{n} \left(\frac{I\{X_i \leq x\} I(\delta_i \epsilon_i = 1)}{1 - G(X_i)} \right) - F_1(x|Z_i) \right\| = o_p(n^{-1/2+r}). \tag{A3}$$

Combining Equations (A2) and (A3), we have

$$\sup_{x,z} \|\hat{F}_1(x|Z) - F_1(x|Z)\| = o_p(n^{-1/2+r})$$

holds uniformly for Z, that is,

$$\|\hat{F}_1 - F_1\|_{\mathcal{H}} \doteq \sup_t \sup_z |\hat{F}_1(t|z) - F_1(t|z)| = o_p(n^{-1/2+r}).$$

□

Lemma A2. *For all positive values $\varepsilon_n = o(1)$, we have*

$$\sup_{\|\beta - \beta_0\| \leq \varepsilon_n, \|F_1^* - F_1\| \leq \varepsilon_n} \|M_n(\beta, F_1^*) - M(\beta, F_1^*) - M_n(\beta_0, F_1)\| = o_p(n^{-1/2}) \tag{A4}$$

Proof. Let Z_{ij} and m_{ij} denote the jth coordinates of Z_i and m_i, respectively. For notational simplicity, in the following we omit the subscript i in various expressions such as Z_i, Z_{ij}, T_i, C_i. Let $K_j, j = 1, \cdots, 5$ be some positive constants. Note that for $j = 1, \cdots, p$,

$$|m_j(\beta, F_1^*) - m_j(\beta', F_1^{*'})|^2 \leq B_1 + B_2 + B_3 + B_4,$$

where

$$\begin{aligned}
B_1 &= Z_j^2 |I\{\epsilon = 1, T \leq C, C \leq g(Z^\top \beta)\} - I\{\epsilon = 1, T \leq C, C \leq g(Z^\top \beta')\}| \\
B_2 &= Z_j^2 |I\{\epsilon = 1, T \leq C, T \leq g(Z^\top \beta), C > g(Z^\top \beta)\} \\
&\quad - I\{\epsilon = 1, T \leq C, T \leq g(Z^\top \beta'), C > g(Z^\top \beta')\}| \\
B_3 &= Z_j^2 \left| \frac{\tau - F_1^*(C)}{1 - F_1^*(C)} \left[I\{F_1^*(C) \leq \tau, C \leq g(Z^\top \beta)\}(1 - I\{T \leq C, \epsilon = 1\}) \right] \right. \\
&\quad \left. - \frac{\tau - F_1^{*'}(C)}{1 - F_1^{*'}(C)} \left[I\{F_1^{*'}(C) \leq \tau, C \leq g(Z^\top \beta')\}(1 - I\{T \leq C, \epsilon = 1\}) \right] \right| \\
B_4 &= Z_j^2 \left| \frac{\tau - F_1^*(C)}{1 - F_1^*(C)} I\{\epsilon = 2, F_1^*(C|Z) \leq \tau, T \leq g(Z^\top \beta), C > g(Z^\top \beta)\} \right. \\
&\quad \left. - \frac{\tau - F_1^{*'}(C)}{1 - F_1^{*'}(C)} I\{\epsilon = 2, F_1^{*'}(C) \leq \tau, T \leq g(Z^\top \beta'), C > g(Z^\top \beta')\} \right|
\end{aligned}$$

It is easy to verify that

$$\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} |I(g(Z^\top\beta) < C) - I(g(Z^\top\beta') < C)|$$
$$\leq \|Z\|\{I(g(Z^\top\beta) - \varepsilon_n < C) - I(g(Z^\top\beta) + \varepsilon_n < C)\}$$

or multiplied by a constant, by Assumption C3. Therefore, by Assumptions A1 and A2,

$$\begin{aligned}
&E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_1\right) \\
&= E\left[\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} Z_j^2 |I\{C \leq g(Z^\top\beta)\} - I\{C \leq g(Z^\top\beta')\}|\right] \\
&\leq E\|Z\|^3\{G(g(Z^\top\beta) + \varepsilon_n) - G(g(Z^\top\beta) - \varepsilon_n)\} \\
&\leq K_1\varepsilon_n.
\end{aligned}$$

Following similar arguments, we can show that

$$\begin{aligned}
&E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_2\right) = E\left[\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} Z_j^2 I(\epsilon_i = 1)\right. \\
&\left.\times|I\{T \leq g(Z^\top\beta), C > g(Z^\top\beta)\} - I\{T \leq g(Z^\top\beta'), C > g(Z^\top\beta')\}|\right] \\
&\leq E[\|Z\|^3\{G(g(Z^\top\beta) + \varepsilon_n) - G(g(Z^\top\beta) - \varepsilon_n)\} \\
&+\|Z\|^3\{F_1(g(Z^\top\beta) + \varepsilon_n) - F_1(g(Z^\top\beta) - \varepsilon_n)\}] \\
&\leq K_2\varepsilon_n.
\end{aligned}$$

Note that

$$\begin{aligned}
B_3 &\leq Z_j^2\left|\left[1 - \frac{1-\tau}{1-F_1^*(C)}\right] I\{F_1^*(C) \leq \tau\} - \left[1 - \frac{1-\tau}{1-F_1^{*'}(C)}\right] I\{F_1^{*'}(C) \leq \tau\}\right| \\
&+Z_j^2\left|I\{C \leq g(Z^\top\beta)\} - I\{C \leq g(Z^\top\beta')\}\right| \\
&\doteq B_{31} + B_{32}.
\end{aligned}$$

Similarly to B_1, it is easy to verify that $E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_{32}\right) \leq K_1\varepsilon_n$. Then

$$\begin{aligned}
B_{31} &= Z_j^2 I\{F_1^*(C) < \tau, F_1^{*'}(C) < \tau\}\frac{(1-\tau)[F_1^*(C) - F_1^{*'}(C)]}{(1-F_1^*(C))(1-F_1^{*'}(C))} \\
&\quad +Z_j^2 I\{F_1^*(C) < \tau < F_1^{*'}(C)\}\frac{1-\tau}{1-F_1^*(C)} \\
&\quad +Z_j^2 I\{F_1^{*'}(C) < \tau < F_1^*(C)\}\frac{1-\tau}{1-F_1^{*'}(C)} \\
&\leq Z_j^2\frac{F_1^*(C) - F_1^{*'}(C)}{(1-\tau)} \\
&\quad +Z_j^2 I\{F_1^*(C) < \tau < F_1^{*'}(C)\} + Z_j^2 I\{F_1^{*'}(C) < \tau < F_1^*(C)\}.
\end{aligned}$$

Since

$$E\left[\sup_{F_1^{*'}:\|F_1^*-F_1^{*'}\|_{\mathcal{H}}} I\{F_1^*(C)<\tau<F_1^{*'}(C)\}\right]$$
$$\leq P\{F_1^*(C)<\tau<F_1^*(C)+\varepsilon_n\}$$
$$\leq G\{F_1^{*-1}(\tau)\}-G\{F_1^{*-1}(\tau-\tau)\}\leq K_3\varepsilon_n.$$

Then by Assumption A1, we have $E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_{31}\right)\leq K_4\varepsilon_n$. Consequently,

$$E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_3\right)\leq K_5\varepsilon_n.$$

Similar arguments to proving B_3, by adding and subtracting $\frac{\tau-F_1^{*'}(C)}{1-F_1^{*'}(C)}I\{\epsilon=2,F_1^{*'}(C)\leq \tau,T\leq g(Z^\top\beta),C>g(Z^\top\beta)\}$, yields

$$\begin{aligned}B_4 &\leq Z_j^2\left|\frac{\tau-F_1^*(C)}{1-F_1^*(C)}I\{F_1^*(C)\leq\tau\}-\frac{\tau-F_1^{*'}(C)}{1-F_1^{*'}(C)}I\{F_1^{*'}(C)\leq\tau\}\right|\\ &\quad+Z_j^2\left|I\{T\leq g(Z^\top\beta'),C>g(Z^\top\beta')\}-I\{T\leq g(Z^\top\beta),C>g(Z^\top\beta)\}\right|\\ &\doteq B_{41}+B_{42}.\end{aligned}$$

By the proof of B_{31} and B_2, we can easily get that $E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_{41}\right)\leq K_4\varepsilon_n$ and $E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_{42}\right)\leq K_2\varepsilon_n$. Thus $E\left(\sup_{\beta':\|\beta-\beta'\|\leq\varepsilon_n} B_4\right)\leq K_5\varepsilon_n$.

Therefore, condition (3.2) of [18] holds with $r=2$ and $s_j=1/2$, and condition (3.3) is satisfied by remark 3(ii) of their paper. Thus, Lemma 2 holds by applying Theorem 3 of [18]. □

Proof of Theorem 1. Note that $F_1(t|Z)<\tau$ is equivalent to $t<g(Z^\top\beta_0)$ and $F_1(g(Z^\top\beta_0)) =\tau$. Therefore, when plugging in the true β_0 and F_1 into M, we get

$$M(\beta_0,F_1)=E\{Z\Big(\tau-H(g(Z^\top\beta_0))-R(\beta_0,F_1)-J(\beta_0,F_1)\Big)\}=0.$$

Because β_0 is the solution of $M(\beta,F_1)$ with $M(\beta,F_1)$ being a continuous function of β in a compact parameter neighborhood $\mathcal{B}$.

Therefore, the consistency of $\hat{\beta}$ is the direct conclusion of Theorem 1 of [18], and we only need verify conditions (1.1), (1.2), and (1.5′) in their paper, as (1.3) is trivially satisfied and (1.4) follows from Lemma A1.

(1.1) By the subgradient condition of quantile regression [13], there exists a vector v with coordinates $|v_i|\leq 1$ such that

$$\|M_n(\hat{\beta},\hat{w})\|=n^{-1}\|(Z_iv_i):i\in\Xi\|=o_p(n^{-1/2}) \tag{A5}$$

by Assumption A.1, where Ξ denotes a $(p+1)-$element subset of $\{1,2,\cdots,n\}$.

(1.2) For any $\varepsilon>0$ and $\beta\in\mathcal{B}$,

$$\begin{aligned}&\inf_{\|\beta-\beta_0\|\geq\varepsilon}\|M(\beta,F_1)\|\\ &=\inf_{\|\beta-\beta_0\|\geq\varepsilon}\|M(\beta,F_1)-\mathrm{M}(\beta_0,F_1)\|\\ &\geq\inf_{\|\beta-\beta_0\|\geq\varepsilon}\|E[ZZ^\top(\beta-\beta_0)]g'(\xi^*)\{1-G(g(\xi^*)|Z)\}f_1(g(\xi^*)|Z)\|,\end{aligned}$$

which is strictly positive under Assumptions A1 and A3. Here ξ^* is some value between $Z^\top\beta$ and $Z^\top\beta_0$.

(1.5′) Let $\{a_n\}$ be a sequence of positive numbers approaching zero as $n \to \infty$. Note that $E\{\|Z_i w_i I(X_i \le g(Z_i^\top\beta))\|^2\} \le E(\|Z_i\|^2) \le K_z$, under Assumption A1. It then follows from Chebyshev's inequality that

$$\sup_{\beta\in\mathcal{B},\|F_1^*-F_1\|_{\mathcal{H}}\le a_n} \|M_n(\beta, F_1^*) - M(\beta, F_1^*)\| = o_p(1).$$

Then the proof of Theorem 1 is complete with the conclusion of Theorem 1 of [18]. □

Proof of Theorem 2. The asymptotic normality of $\hat{\beta}$ relies on the results of Theorem 2 in [18]. We need to prove conditions (2.1)–(2.4), (2.5′), and (2.6′) in their paper. Conditions (2.1), (2.4), and (2.5′) hold directly by (A5), Lemma A1, and Lemma A2, respectively.

Note that for any C_i lying above the τth conditional quantile $Z_i^\top\beta_0$, the quantile fit will not be affected if we assign the entire weight to either (Z_i, C_i) or $(Z_i, X^{+\infty})$. Then we obtain

$$\begin{aligned}
\mathbf{\Gamma}_1(\beta_0, F_1) &= \frac{\partial M(\beta, F_1)}{\partial\beta}|_{\beta=\beta_0} \\
&= -E[ZZ^\top g'(Z^\top\beta_0)\{1 - G(g(Z^\top\beta_0)|Z)\}f_1(g(Z^\top\beta_0)|Z)],
\end{aligned}$$

which is continuous at β_0 and of full rank under Assumption A3. For all $\beta \in \mathcal{B}$, we define the functional derivative of $M(\beta, F_1^*)$ at F_1 in the direction $[F_1^* - F_1]$ as

$$\begin{aligned}
&\mathbf{\Gamma}_2(\beta, F_1)[F_1^* - F_1] \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}[M\{\beta, F_1 + \varepsilon(F_1^* - F_1)\} - M\{\beta, F_1\}] \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}E[R(\beta, F_1) - R(\beta, F_{1\varepsilon}) + J(\beta, F_1) - J(\beta, F_{1\varepsilon})]
\end{aligned}$$

where $F_{1\varepsilon} = F_1 + \varepsilon(F_1^* - F_1)$. Since

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}EZ[R(\beta, F_1) - R(\beta, F_{1\varepsilon})] \\
&= EZ[A_1(\beta) + A_2(\beta)] + \\
&(1-\tau)EZ\int_0^{g(Z^\top\beta)} g_0(u)I\{F_1(u|Z) \le \tau\}\frac{F_1^*(u|Z) - F_1(u|Z)}{1 - F_1(u|Z)}du
\end{aligned}$$

where

$$\begin{aligned}
A_1(\beta) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^{g(Z^\top\beta)} g_0(u)(1 - F_1(u))[I\{F_1(u|Z) \le \tau\} - I\{F_{1\varepsilon}(u|Z) \le \tau\}du] \\
A_2(\beta) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^{g(Z^\top\beta)} g_0(u)(1 - F_1(u))(1-\tau)\frac{I\{F_{1\varepsilon}(u|Z) \le \tau\} - I\{F_1(u|Z) \le \tau\}}{1 - F_{1\varepsilon}(u|Z)}du.
\end{aligned}$$

Similarly , we can derive

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}EZ[J(\beta, F_1) - J(\beta, F_{1\varepsilon})] \\
&= EZ[A_3(\beta) + A_4(\beta)] + (1-\tau)EZ[F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \\
&\times \int_{g(Z^\top\beta)}^{\infty} g_0(u)I\{F_1(u|Z) \le \tau\}\frac{F_1^*(u|Z) - F_1(u|Z)}{(1 - F_1(u|Z))^2}du
\end{aligned}$$

where

$$\begin{aligned}
A_3(\beta) &= [F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{g(Z^\top\beta)}^{\infty} g_0(u) \\
&\quad I\{F_1(u|Z) \le \tau\} - I\{F_{1\varepsilon}(u|Z) \le \tau\} du \\
A_4(\beta) &= [F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)](1-\tau) \\
&\quad \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{g(Z^\top\beta)}^{\infty} g_0(u)\frac{I\{F_{1\varepsilon}(u|Z) \le \tau\} - I\{F_1(u|Z) \le \tau\}}{1-F_{1\varepsilon}(u|Z)} du.
\end{aligned}$$

For β such that $g(Z^\top\beta) < g(Z^\top\beta_0)$, $A_1(\beta) = 0, A_2(\beta) = 0$. For sufficiently small ε, $F_{1\varepsilon}^{-1}(\tau) > g(Z^\top\beta)$, then

$$\begin{aligned}
A_3(\beta) &= [F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{G(g(Z^\top\beta_0)) - G(F_{1\varepsilon}^{-1}(\tau|Z))\Big\} \\
A_4(\beta) &= (1-\tau)[F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{\tilde{G}(F_{1\varepsilon}^{-1}(\tau|Z)) - \tilde{G}(g(Z^\top\beta_0|Z))\Big\}
\end{aligned}$$

where $\frac{d\tilde{G}(u|Z)}{du} = \frac{g_0(u)}{1-F_1(u|Z)}$.

For β such that $g(Z^\top\beta) > g(Z^\top\beta_0)$, $A_3(\beta) = 0, A_4(\beta) = 0$. For sufficiently small ε, $F_{1\varepsilon}^{-1}(\tau) < g(Z^\top\beta)$, then

$$A_1(\beta) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{\check{G}(g(Z^\top\beta_0|Z)) - \check{G}(F_{1\varepsilon}^{-1}(\tau|Z))\Big\}$$

where $\frac{d\check{G}(u|Z)}{du} = g_0(u)(1-F_1(u|Z))$ and

$$A_2(\beta) = (1-\tau)\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{G(F_{1\varepsilon}^{-1}(\tau|Z)) - G(g(Z^\top\beta_0))\Big\}.$$

For $\beta = \beta_0$, note that $I\{F_0(t|Z) < \tau\} = 1$ for $t \in (0, g(Z^\top\beta))$, then

$$\begin{aligned}
A_1(\beta) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{\check{G}(g(Z^\top\beta_0|Z)) - \check{G}(F_{1\varepsilon}^{-1}(\tau|Z))\Big\} \\
A_2(\beta) &= (1-\tau)\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{G(F_{1\varepsilon}^{-1}(\tau|Z)) - G(g(Z^\top\beta_0))\Big\}
\end{aligned}$$

and $I\{F_0(t|Z) < \tau\} = 0$ for $t \in (g(Z^\top\beta), \infty)$, then

$$\begin{aligned}
A_3(\beta) &= [F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{G(g(Z^\top\beta_0)) - G(F_{1\varepsilon}^{-1}(\tau|Z))\Big\} \\
A_4(\beta) &= (1-\tau)[F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \\
&\quad \cdot \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big\{\tilde{G}(F_{1\varepsilon}^{-1}(\tau|Z)) - \tilde{G}(g(Z^\top\beta_0|Z))\Big\}.
\end{aligned}$$

By expanding $\tilde{G}(F_{1\varepsilon}^{-1}(\tau|Z))$ (treated as a function of ε) around $\varepsilon = 0$, and using the fact that $\frac{d}{d\varepsilon}F_{1\varepsilon}^{-1}(\tau|Z)|_{\varepsilon=0} = \frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))}$ (example 20.5 in [19]), we obtain

$$G(F_{1\varepsilon}^{-1}(\tau|Z)) = G(g(Z^\top\beta_0)) + g_0(g(Z^\top\beta_0))\frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))}\varepsilon + O(\varepsilon^2).$$

Similarly, we have

$$\begin{aligned}\breve{G}(F_{1\varepsilon}^{-1}(\tau|Z)) &= \breve{G}(g(Z^\top\beta_0)) + g_0(g(Z^\top\beta_0))(1 - F_1(g(Z^\top\beta_0))) \\ &\quad \cdot \frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))}\varepsilon + O(\varepsilon^2). \\ \tilde{G}(F_{1\varepsilon}^{-1}(\tau|Z)) &= \tilde{G}(g(Z^\top\beta_0)) + \frac{g_0(g(Z^\top\beta_0))}{1 - F_1(g(Z^\top\beta_0))} \\ &\quad \cdot \frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))}\varepsilon + O(\varepsilon^2).\end{aligned}$$

Therefore, for β such that $g(Z^\top\beta) < g(Z^\top\beta_0)$,

$$\begin{aligned}A_3(\beta) + A_4(\beta) &= -[F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)]g_0(g(Z^\top\beta_0)) \\ &\quad \frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))}\frac{F_1(g(Z^\top\beta_0)) - \tau}{1 - F_1(g(Z^\top\beta_0))} \equiv 0\end{aligned}$$

for β such that $g(Z^\top\beta) \geq g(Z^\top\beta_0)$,

$$A_1(\beta) + A_2(\beta) = g_0(g(Z^\top\beta_0))(F_1(g(Z^\top\beta_0)) - \tau)\frac{\tau - F_1^*(g(Z^\top\beta_0))}{f_1(g(Z^\top\beta_0))} \equiv 0.$$

That is

$$\begin{aligned}&\boldsymbol{\Gamma}_2(\beta_0, F_1)[F_1^* - F_1] \\ &= (1-\tau)EZ\int_0^{g(Z^\top\beta)} g_0(u)I\{F_1(u|Z) \leq \tau\}\frac{F_1^*(u|Z) - F_1(u|Z)}{1 - F_1(u|Z)}du \\ &\quad +(1-\tau)EZ[F_0(g(Z^\top\beta)|Z) - F_1(g(Z^\top\beta)|Z)] \\ &\quad\quad \times \int_{g(Z^\top\beta)}^{\infty} g_0(u)I\{F_1(u|Z) \leq \tau\}\frac{F_1^*(u|Z) - F_1(u|Z)}{(1 - F_1(u|Z))^2}du. \end{aligned} \tag{A6}$$

With the process of Taylor expansion, we can verify condition (2.3) of [18] under Assumptions A1 and A2.

Then, we verify condition (2.6). Combining (A6) and the analysis above, we have

$$\begin{aligned}&\boldsymbol{\Gamma}_2(\beta_0, F_1)[\hat{F}_1 - F_1] \\ &= (1-\tau)EZ\int_0^{g(Z^\top\beta_0)} g_0(u)\frac{\hat{F}_1(u|Z) - F_1(u|Z)}{1 - F_1(u|Z)}du\end{aligned} \tag{A7}$$

Denote $F_1^G(t|Z) = \frac{1}{n}\sum_{i=1}^n \frac{I\{X_i \leq t, \delta_i\epsilon_i = 1\}}{1 - G(X_i)}$, $N_i^G(t) = I(X_i \leq t, \delta_i\epsilon_i = 0)$, $Y_i(t) = I(X_i \geq t)$, $y(t) = P(X \geq t)$, $\lambda^G(t) = \lim_{\Delta\to 0} P(X \in (t, t+\Delta)|X \geq t)$, $\Lambda^G(t) = \int_0^t \lambda^G(s)ds$ and $M_i^G(t) = N_i^G - \int_0^\infty Y_i(s)d\Lambda^G(s)$. Follow the proof in [5], $\sup_{t\in[0,\nu)} \|n^{1/2}\{\hat{G}(t) - G(t) - n^{-1/2}\sum_{i=1}^n G(t)\int_0^t y(s)^{-1}dM_i^G\}\| \to 0$, from [17], and $n^{-1}\sum_{i=1}^n Y_i(t)I\{X_i \leq x\}I(\delta_i\epsilon_i = 1)(1 - G(X_i))^{-1}$ converges to $\pi(x,t)$ uniformly in both $x \in R$ and $t \in [0,\nu)$, where $\pi(x,t) = EY_i(t)I\{X_i \leq x\}I(\delta_i\epsilon_i = 1)(1 - G(X_i))^{-1}$. Then

$$
\begin{aligned}
&\hat{F}_1(x|Z) - F_1(x|Z) = F_1^G(x|Z) - F_1(x|Z) + \hat{F}_1(x|Z) - F_1^G(x|Z) \\
= &\ \frac{1}{n}\sum_{i=1}^n \xi_{1,i}(x) - \frac{1}{n}\sum_{i=1}^n \frac{\hat{G}(X_i) - G(X_i)}{\hat{G}(X_i)G(X_i)} I(X_i \le x) I(\delta_i \epsilon_i = 1) \\
\approx &\ \frac{1}{n}\sum_{i=1}^n \xi_{1,i}(x) - \frac{1}{n}\sum_{i=1}^n \frac{n^{-1}\sum_{j=1}^n Y_i(s) y(s)^{-1} dM_j^G}{G(X_i)} I(X_i \le x) I(\delta_i \epsilon_i = 1) \\
= &\ \frac{1}{n}\sum_{i=1}^n \xi_{1,i}(x) - \frac{1}{n}\sum_{i=1}^n \int_0^\infty \left(\sum_{j=1}^n \frac{Y_j(s) I(X_j \le x) I(\delta_j \epsilon_j = 1)}{nG(X_j)} \right) \frac{dM_i^G(s)}{y(s)} \\
\approx &\ \frac{1}{n}\sum_{i=1}^n \xi_{1,i}(x) - \frac{1}{n}\sum_{i=1}^n \int_0^\infty \pi(x,s) \frac{dM_i^G(s)}{y(s)} \\
= &\ \frac{1}{n}\sum_{i=1}^n \{\xi_{1,i}(x) - \xi_{2,i}(x)\},
\end{aligned}
$$

where $\approx$ denotes asymptotic equivalence uniformly in $\tau \in [\tau_L, \tau_U]$, $\xi_{1,i}(x) = I(X_i \le x) I(\delta_i \epsilon_i = 1) G(X_i)^{-1} - F_1(x|Z)$ and $\xi_{2,i} = \int_0^\infty \pi(x,s) y(s)^{-1} dM_i^G(s), i = 1, \ldots, n$. Similarly derived as [5], $\int_0^\infty \pi(x,s) y(s)^{-1} dM_i^G$ is Lipshitz in x, $\hat{F}_1(x|Z) - F_1(x|Z)$ converges weakly to a mean zero Guassian process with covariance matrix $\Sigma(x) = E\{\xi_1(x)'\xi_1(x)\}$. Then by (A7),

$$
\begin{aligned}
&\mathbf{\Gamma}_2(\beta_0, F_1)[\hat{F}_1 - F_1] \\
\approx &\ (1-\tau) n^{-1} \sum_{i=1}^n E_z \left[Z \int_0^{g(Z^\top \beta_0)} g_0(u) \frac{\xi_{1,i}(u) - \xi_{2,i}(u)}{1 - F_1(u|Z)} du \right] \\
= &\ (1-\tau) n^{-1} \sum_{i=1}^n \phi_i
\end{aligned} \tag{A8}
$$

where

$$
\phi_i = E_z Z \int_0^{g(Z^\top \beta_0)} g_0(u) \frac{\xi_{1,i}(u) - \xi_{2,i}(u)}{1 - F_1(u|Z)} du \tag{A9}
$$

is a random vector with mean 0 and $E\|\phi_i\|^2 < \infty$ by Assumptions A1–A3.

Recall $M_n(\beta_0, F_1) = n^{-1}\sum_{i=1}^n m_i(\beta_0, F_1)$ being independent mean 0 random vectors.

$$
\begin{aligned}
&m_i(\beta_0, F_1) \\
= &\ Z_i \Big(\tau - I\{\epsilon_i = 1, T_i \le C_i, C_i \le g(Z_i^\top \beta_0)\} \\
&- I\{\epsilon_i = 1, T_i \le C_i, g^{-1}(T_i) \le Z_i^\top \beta_0, C_i > g(Z_i^\top \beta_0)\} \\
&- \frac{\tau - F_1(C_i)}{1 - F_1(C_i)} \Big[I\{F_1(C_i) \le \tau, C_i \le g(Z_i^\top \beta_0)\}(1 - I\{T_i \le C_i, \epsilon_i = 1\}) \Big] \Big) \\
\doteq &\ Z_i(\tau - D_1 - D_2 - D_3).
\end{aligned}
$$

Since $Em_i(\beta_0, F_1) = 0$, and $D_i D_j = 0$ for $i \ne j$, it is easy to verify

$$
\begin{aligned}
&\mathrm{Cov}\{m_i(\beta_0, F_1)\} \\
&= E_{Z,C} E \left\{ Z_i Z_i^\top \left[\tau(1-\tau) I(C_i > g(Z_i^\top \beta_0)) + I(C_i \le g(Z_i^\top \beta_0)) \frac{F_1(C_i)(1-\tau)^2}{1 - F_1(C_i)} \right] \right\} \doteq d_1.
\end{aligned}
$$

Then applying the central limit theorem gives

$$
n^{1/2}\{M_n(\beta_0, F_1) + \mathbf{\Gamma}_2(\beta_0, F_1)[\hat{F}_1 - F_1]\} \xrightarrow{D} N(0, \mathbf{V}),
$$

where

$$\begin{aligned}\mathbf{V} &= \mathrm{Cov}\{m_i(\beta_0, F_1) + (1-\tau)\phi_i\} \doteq d_1 + d_2 + d_2 \\ d_1 &= (1-\tau)E\{m_i(\beta_0, F_1)\phi^\top\} \\ d_2 &= (1-\tau)^2 E\{\phi^\top \phi\}\end{aligned}$$

Then the proof for (14) is thus complete by Theorem 2 of [18]. □

Proof of Theorem 3. Let $\hat{\mathcal{A}}_n = \{j : \tilde{\beta}_j \neq 0\}$. We first show that for any $j \neq \mathcal{A}$, $P(j \in \hat{\mathcal{A}}_n) \to 0$ as $n \to \infty$. Suppose there exists a $k \in \hat{\mathcal{A}}^c$ such that $|\tilde{\beta}_k| \neq 0$. Let β^* be a vector constructed by replacing $\tilde{\beta}_k$ with 0 in $\tilde{\beta}$. For simplicity, we write $\hat{w}_i = w_i(\hat{F}_1)$. Note that $|\rho_\tau(a) - \rho_\tau(b)| \leq |a-b| \max\{\tau, 1-\tau\} < |a-b|$. Therefore, for large enough n,

$$\begin{aligned}&Q_p(\tilde{\beta}, \hat{w}_i) - Q_p(\beta^*, \hat{w}_i) \\ &= \sum_{i=1}^n \hat{w}_i \Big\{\rho_\tau(g^{-1}(X_i) - Z_i^\top \tilde{\beta}) - \rho_\tau(g^{-1}(X_i) - Z_i^\top \beta^*)\Big\} \\ &+ \sum_{i=1}^n (1-\hat{w}_i)\Big\{\rho_\tau(g^{-1}(X^{+\infty}) - Z_i^\top \tilde{\beta}) - \rho_\tau(g^{-1}(X^{+\infty}) - Z_i^\top \beta^*)\Big\} + p_{\lambda_n}(|\tilde{\beta}_k|) \\ &\geq -2\sum_{i=1}^n \|Z_i\| \cdot |\hat{\beta}_k| + \lambda_n |\hat{\beta}_k|^{-\gamma} |\tilde{\beta}_k|.\end{aligned}$$

By Theorem 1, $\hat{\beta}_k - \beta_k = O_p(n^{-1/2})$ and β_k, thus $\hat{\beta}_k = O_p(n^{-1/2})$. As $\sum_{i=1}^n \|Z_i\| = O_p(1)$ and $n^{-1}\lambda_n |\hat{\beta}_k|^{-\gamma} \geq n^{r/\gamma - 1}\lambda_n \to \infty$, which yields

$$\begin{aligned}&Q_p(\tilde{\beta}, \hat{w}_i) - Q_p(\beta^*, \hat{w}_i) \geq -2\sum_{i=1}^n \|Z_i\| \cdot |\hat{\beta}_k| + \lambda_n |\hat{\beta}_k|^{-\gamma} |\tilde{\beta}_k| \\ &\geq |\tilde{\beta}_k| n\Big[-O_p(1) + n^{-1}\lambda_n |\hat{\beta}_k|^{-\gamma}\Big] \geq c^* n^{\gamma/2 - 1/2}\lambda_n > 0, \text{as} \quad n \to \infty,\end{aligned} \tag{A10}$$

where c^* is any positive constant. This contradicts the fact that $Q_p(\tilde{\beta}, \hat{w}_i) \leq Q_p(\beta^*, \hat{w}_i)$.

We next show that for any $j \in \mathcal{A}$, $P(j \notin \hat{\mathcal{A}}_n) \to 0$. We write $b_{\mathcal{A}} = (b_j, j \in \mathcal{A})$ for any vector $b \in R^p$, and $B_{\mathcal{A}\mathcal{A}}$ as the sub-matrix of a $(p+1) \times (p+1)$ matrix B with both row and column indices in $\mathcal{A}$. By Taylor expansion

$$\begin{aligned}M_n(\beta_{\mathcal{A}}, F_1^*) &= M_n(\beta_{0\mathcal{A}}, F_1) + \Gamma_{1\mathcal{A}\mathcal{A}}(\beta_{\mathcal{A}} - \beta_{0\mathcal{A}}) \\ &+ \Gamma_{2\mathcal{A}\mathcal{A}}(\beta_{0\mathcal{A}}, F_1)[F_1^* - F_1] + o_p(n^{-1/2})\end{aligned} \tag{A11}$$

uniformly over $\beta_{\mathcal{A}}, F_1$ such that $\|\beta_{\mathcal{A}} - \beta_{0\mathcal{A}}\| = O(n^{-1/2})$ and $\|F_1^* - F_1\|_{\mathcal{H}} = o(n^{-1/2+r})$. Let $\beta_{\mathcal{A}} - \beta_{0\mathcal{A}} = n^{-1/2}u$, we have

$$\begin{aligned}nu^\top M_n(\beta_{\mathcal{A}}, \hat{F}_1) &= nu^\top\{M_n(\beta_{0\mathcal{A}}, F_1) + \Gamma_{2\mathcal{A}\mathcal{A}}\} \\ &+ n^{1/2}u^\top \Gamma_{1\mathcal{A}\mathcal{A}} u + o_p(n^{1/2})\end{aligned} \tag{A12}$$

where $\Gamma_{2\mathcal{A}\mathcal{A}} = \Gamma_{2\mathcal{A}\mathcal{A}}(\beta_{0\mathcal{A}}, F_1)[\hat{F}_1 - F_1]$. Therefore, with probability tending to 1,

$$\begin{aligned}-nu^\top M_n(\beta_{\mathcal{A}}, \hat{F}_1) &\geq -nu^\top\{M_n(\beta_{0\mathcal{A}}, F_1) + \Gamma_2\} - n^{1/2}u^\top \Gamma_{1\mathcal{A}\mathcal{A}} u + o(n^{1/2}) \\ &\geq k_0 n^{1/2+r}\end{aligned} \tag{A13}$$

for some positive k_0 and $r > 0$. However, the subgradient condition (A5) requires that

$$\|nu^\top M_n(\beta_{\mathcal{A}}, \hat{F}_1)\| + \lambda_n \sum_{j \in \mathcal{A}} |\hat{\beta}_j|^{-r} |\tau - I(\tilde{\beta}_j < 0)| \leq O_p(\max_i \|Z_i\|). \tag{A14}$$

When $\lambda_n = o(n^{1/2})$ and Assumption A1 holds, (A13) and (A14) suggest that the subgradient condition cannot hold if $\|\tilde{\beta}_{\mathcal{A}} - \beta_{0\mathcal{A}}\| = Kn^{-1/2}$ for some positive K. Using the monotonicity argument in [20] , we can show that the subgradient condition also cannot hold if $\|\tilde{\beta}_{\mathcal{A}} - \beta_{0\mathcal{A}}\| > Kn^{-1/2}$. Therefore, $\|\tilde{\beta}_{\mathcal{A}} - \beta_{0\mathcal{A}}\| \leq Kn^{-1/2}$ with probability tending to 1. Equivalently speaking, for all $j \in \mathcal{A}$, $P(j \in \hat{\mathcal{A}}_n) \to 1$ or $P(j \notin \hat{\mathcal{A}}_n) \to 0$. The proof of Theorem 3 is thus complete. □

References

1. Scrucca, L.; Santucci, A.; Aversa, F. Regression modeling of competing risk using R: An in depth guide for clinicians. *Bone Marrow Transpl.* **2010**, *45*, 1388–1395. [CrossRef] [PubMed]
2. Fine, J.P. ; Gray, R.J. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **1999**, *94*, 496–509. [CrossRef]
3. Fu, Z.; Parikh, C.R.; Zhou, B.J. Penalized variable selection in competing risks regression. *Lifetime Data Anal.* **2017**, *23*, 353–376. [CrossRef] [PubMed]
4. Koenker, R.W.; Bassett, G. Regression quantile. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
5. Peng, L.; Fine, J.P. Competing risks quantile regression. *J. Am. Stat. Assoc.* **2009**, *104*, 1440–1453. [CrossRef]
6. Sun, Y.Q.; Wang, H.; Gilbert, P. Quantile regression for competing risks data with missing cause of failure. *Stat. Sin.* **2012**, *22*, 703–728. [CrossRef] [PubMed]
7. Ahn, K.W.; Kim, S. Variable selection with group structure in competing risks quantile regression. *Stat. Med.* **2018**, *37*, 1577–1586. [CrossRef] [PubMed]
8. Li, E.; Tian, M.; Tang, M. Variable selection in competing risks models based on quantile regression. *Stat. Med.* **2019**, *38*, 4670–4685. [CrossRef] [PubMed]
9. Wang, H.J. ; Wang, L. Locally weighted censored quantile regression. *J. Am. Stat. Assoc.* **2009**, *104*, 1117–1128. [CrossRef]
10. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B.* **1996**, *58*, 267–288. [CrossRef]
11. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 121–152. [CrossRef]
12. Gray, R.J. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Stat.* **1988**, *16*, 1141–1154. [CrossRef]
13. Koenker, R. *Quantile Regression*; Cambridge University Press: New York, NY, USA, 2005.
14. Wang, H.J.; Zhou, J.; Li, Y. Variable selection for censored quantile regresion. *Stat. Sin.* **2013**, *23*, 145–167. [PubMed]
15. Robins, J.M. ; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology Theorethodological Issues*; Jewell, N., Dietz, K., Farewell, V., Eds.; Birkhäuser: Boston, MA, USA, 1992; pp. 24–33.
16. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [CrossRef]
17. Pepe, M.S. Inference for Events With Dependent Risks in Multiple Endpoint Studies. *J. Am. Stat. Assoc.* **1991**, *86*, 770–778. [CrossRef]
18. Chen, X.; Linton, O.; Van Keilegom, I. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **2003**, *71*, 1591–1608. [CrossRef]
19. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998.
20. Jureckova, J. Asymptotic relations of m-estimates and r-estimates in linear regression. *Ann. Statist.* **1977**, *5*, 464–472. [CrossRef]