CrossMark

Accelerated failure time model with quantile information

Mu Zhao[1] · Yixin Wang[2] · Yong Zhou[2,3]

Abstract The censored linear regression model, also referred to as the accelerated failure time model, is a useful alternative to the popular Cox model in the analysis of censored survival data. In this paper, we combine the quantile information with censored least-squares normal equations to get estimators with smaller estimated standard error for regression parameters. An inverse probability-weighted method is proposed to construct unbiased estimating equations with censored data and the lack of smoothness of the objective equations is overcome by replacing them with smooth approximations. The proposed estimators are established based on the empirical likelihood method and generalized method of moments, respectively, and their asymptotic properties are stud-

✉ Yixin Wang
wangyixin0812@163.com

Mu Zhao
zhaomu266@163.com

Yong Zhou
yzhou@amss.ac.cn

[1] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China

[2] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

[3] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

ied under some regular conditions. We also conduct some simulation experiments to investigate the finite-sample properties of the proposed estimators. The Stanford Heart Transplant data are used to illustrate the proposed estimating method.

1 Introduction

The least-squares approach for fitting a linear regression model provides a statistical technique for investigating the relationship between variables. Its simple structure and ease of interpretation have made it an attractive method for practitioners. Another important approach for fitting a linear model is the quantile regression method originated by Koenker and Bassett (1978). A collection of conditional quantiles can characterize the entire conditional distribution and capture the rich underlying relationship between the quantiles of response variable and covariates. But a major difficulty of quantile regression is that to obtain the asymptotic covariance matrix of estimators, we need an estimation of the regressor density, which is often cumbersome to obtain. This motivated Zhou et al. (2011) to bring together these two well-known techniques, and develop a coherent estimation framework that can be applied to a myriad of situations. Their simulation studies and real data analysis have shown that the least-squares estimator combined with auxiliary quantile information not only leads to a more efficient estimator, but also results in a relatively simple calculation of estimator's standard error that does not require any density estimation. Based on the same idea, Liu and Ishfaq (2011) and Liu et al. (2011) considered the estimation of distribution function when auxiliary quantile information is available with complete data and missing data, respectively. The aim of this paper is to extend this idea to censored linear regression model.

The censored linear regression model, also referred to as the accelerated failure time (AFT) model, specifies that the logarithm of the failure time T is related to a $p \times 1$ vector of covariate Z_i in the following way:

$$\log T_i = Z_i^{\tau} \beta_0 + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where β_0 is a $p \times 1$ vector of unknown regression parameters and ϵ_i, $i = 1, \ldots, n$, are independent and identically distributed with an unspecified common distribution function F_{ϵ}, but zero mean and finite variance. This model, as an alternative to the popular Cox model, has been studied extensively in literature, see, for instance, Buckley and James (1979), Koul et al. (1981), Lai and Ying (1991), Ritov (1990) and Wei (1992), among others. When the data are completely observed, Zhou et al. (2011) suggested to estimate β_0 in (1) based on the following estimating function:

$$\psi(T, Z, \beta) = \begin{pmatrix} \psi_{(1)}(T, Z, \beta) \\ \psi_{(2)}(T, Z, \beta) \end{pmatrix} = \begin{pmatrix} Z(\log T - Z^{\tau}\beta) \\ Z(\frac{1}{2} - I(\log T - Z^{\tau}\beta \le 0)) \end{pmatrix}. \tag{2}$$

The first part of (2) is based on the normal equation of least squares and the second part of (2) is based on the auxiliary quantile information, using the assumption that

the errors are symmetric or the median of the errors is zero. In general, if information corresponding to the ζ-th quantile is known, then $1/2$ in (2) may be replaced by ζ. In survival analysis, the survival time T is usually right censored by another variable C. The observed data are (X_i, Δ_i, Z_i), $i = 1, 2, \ldots, n$, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, and $I(\cdot)$ is the indicator function. Assume that T_i and C_i have absolutely continuous survival function $S(t)$ and $K(t)$, respectively, and C_i is independent of T_i and Z_i. In this paper, we will extend the results of Zhou et al. (2011) to right censored data case.

The rest of this paper is arranged as follows. A smoothing technique is introduced in Sect. 2. In Sect. 3, we propose an IPW estimating equation method to construct asymptotic unbiased estimating equations with right censored data. In Sect. 4, we propose two estimators based on EL and GMM methods, respectively, and we also show that the proposed two estimators are asymptotically normally distributed. Section 5 reports some simulation results and a real data example. A discussion is given in Sect. 6, and the proofs of the theorems are contained in the Appendix.

2 A smoothing technique for non-smooth EEs

Note that based on (2), we can construct $2p$ estimating equations (EEs), but we only have p unknown parameters. This is so-called over-determined case. Obviously, ordinary estimating methods are infeasible here. The common procedures equipped to handle the over-determined case are the generalized method of moments (GMM) given by Hansem (1982) and empirical likelihood (EL) method developed by Qin and Lawless (1994). However, the functions obtained from quantile information are non-differentiable in β, since they are indicators.

The lack of smoothness of the objective equations can be handled by replacing them with smooth approximations because smoothness of the objective function is required for Taylor expansions, such as Chen and Hall (1993), Heller (2007) and Song et al. (2007). Similar to the smoothing technique developed in Zhou et al. (2011), the proposed smoothed estimating equations in this paper are also kernel based. Without loss of generality, we illustrate the idea by assuming $p = 1$. Here, we use the high-order kernel to smooth the estimating equations associated with quantiles as those in existing literature, for more details please refer to Zhou and Jing (2003). Consider a smooth kernel function $l(\xi)$, then $L(t) = \int_{-\infty}^{t} l(u) \mathrm{d}u$ is also a smooth function, and $b_n \to 0$, $nb_n \to \infty$ as $n \to \infty$. In practice, $l(\xi)$ can be any smooth function, for example, $l(\xi)$ may be an r-th order kernel function such that

$$\int u^j l(u) \mathrm{d}u = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r - 1, \\ c_r \neq 0, & \text{if } j = r, \end{cases}$$

for some integer $r \geq 2$. The smoothed version of the second part of (2) comes out to be

$$\phi_{(2)}(T, Z, \beta) = Z \left\{ \frac{1}{2} - \int I(\log T - Z^{\tau}\beta + b_n \xi \le 0) dL(\xi) \right\}$$
$$= Z \left\{ \frac{1}{2} - L \left(\frac{Z^{\tau}\beta - \log T}{b_n} \right) \right\}. \quad (3)$$

From Zhou et al. (2011), we know that

$$E\phi_{(2)}(T, Z, \beta_0) = E \left\{ Z \left[\frac{1}{2} - \int I(\log T - Z^{\tau}\beta_0 + b_n \xi \le 0) dL(\xi) \right] \right\}$$
$$\xrightarrow{b_n \to 0} E\psi_{(2)}(T, Z, \beta_0) = 0. \quad (4)$$

3 Inverse probability-weighted EEs with censored data

With complete data, we can get the estimators of regression parameters along lines of Zhou et al. (2011) based on the smoothed estimating equations

$$\sum_{i=1}^{n} \phi(X_i, Z_i, \beta) = \sum_{i=1}^{n} \begin{pmatrix} \psi_{(1)}(X_i, Z_i, \beta) \\ \phi_{(2)}(X_i, Z_i, \beta) \end{pmatrix} = 0. \quad (5)$$

While, in the presence of censoring, (5) is no longer asymptotically unbiased estimating equations. Hence, we consider the modified estimating equations which are called inverse probability-weighted (IPW) estimating equations,

$$\sum_{i=1}^{n} \frac{\Delta_i}{K(X_i)} \phi(X_i, Z_i, \beta) = 0. \quad (6)$$

The above idea of weighting the complete observations by their inverse probabilities was originated by Horvitz and Thompson (1952) in the context of sample surveys. The adaptation of this idea to the setting of censored survival data was initially considered by Koul et al. (1981), and later on by Robins and Rotnitzky (1992) and Lin and Ying (1993). Zhao and Tsiatis (1997) applied this idea to the problem of quality adjusted survival time. Recently, Bang and Tsiatis (2000), Lin (2000) and Bang and Tsiatis (2002) used this method to estimate medical costs. We find that (6) is an asymptotically unbiased estimating equation which is a consequence of the following equality:

$$E \left\{ \frac{\Delta}{K(X)} \phi(X, Z, \beta) \right\} = E \left\{ E(\Delta|T, Z) \frac{1}{K(T)} \phi(T, Z, \beta) \right\}$$
$$= E\{\phi(T, Z, \beta)\} \approx 0.$$

In practice, the survival function $K(\cdot)$ is unknown. Here, we propose to estimate $K(\cdot)$ by the Kaplan–Meier estimator (Kaplan and Meier 1958) with the roles of censoring time C_i and survival time T_i reversed. That is,

$$\widehat{K}(t) = \Pi_{u \leq t} \left[1 - \frac{\mathrm{d}N^c(u)}{Y(u)} \right],$$

where $N^c(u) = \sum_{i=1}^{n} I(X_i \leq u, \Delta_i = 0)$, $Y(u) = \sum_{i=1}^{n} I(X_i \geq u)$. The simple weighted complete-case EEs come out to be:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i)} \phi(X_i, Z_i, \beta) = 0. \tag{7}$$

In the next section, we give the estimators of regression parameters based on the IPW estimating equations (7) by empirical likelihood and GMM methods.

4 Inference based on IPW EEs

4.1 Empirical likelihood

In this section, we construct an estimated empirical likelihood to make statistical inference on β. For convenience, denote $\Delta_i / \widehat{K}(X_i) = V_{ni}$. Let $p = (p_1, \ldots, p_n)$, $p_i \geq 0$ for all $1 \leq i \leq n$ with $\sum_{i=1}^{n} p_i = 1$. Define F_p to be the distribution function which assigns probability p_i to the point $V_{ni} \phi(X_i, Z_i, \beta)$. The empirical likelihood is

$$L(\beta) = \Pi_{i=1}^{n} p_i. \tag{8}$$

We maximize (8) subject to restrictions

$$p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} V_{ni} \phi(X_i, Z_i, \beta) p_i = 0. \tag{9}$$

For any given β, let the set $\Omega_\beta = \{\lambda : 1 + \lambda^\tau V_{ni} \phi(X_i, Z_i, \beta) \geq 1/n\}$ be convex, closed and also bounded if the convex hull of $V_{ni} \phi(X_i, Z_i, \beta)$ contains 0. By the Lagrange multiplier method, we have

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^\tau V_{ni} \phi(X_i, Z_i, \beta)},$$

where λ is the solution to

$$\frac{1}{n} \sum_{i=1}^{n} \frac{V_{ni} \phi(X_i, Z_i, \beta)}{1 + \lambda^\tau V_{ni} \phi(X_i, Z_i, \beta)} = 0. \tag{10}$$

Note that $\Pi_{i=1}^{n} p_i$ subject to $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0, 1 \leq i \leq n$, attains its maximum value n^{-n} at $p_i = n^{-1}$. Hence, we define the profile empirical likelihood ratio by

$$R(\beta) = \Pi_{i=1}^{n} (np_i) = \Pi_{i=1}^{n} \frac{1}{1 + \lambda^\tau V_{ni} \phi(X_i, Z_i, \beta)}.$$

The log empirical likelihood ratio multiplied by -2 is then given by

$$\mathcal{R}(\beta) = -2\log R(\beta) = 2\sum_{i=1}^{n}\log\{1 + \lambda^{\tau}V_{ni}\phi(X_i, Z_i, \beta)\}. \tag{11}$$

Let $\widehat{\beta}_e$ be the MELE that results from minimizing $\mathcal{R}(\beta)$, then we have:

Theorem 1 *Let Assumptions 1–6 in the Appendix be satisfied. Then,*

$$\sqrt{n}(\widehat{\beta}_e - \beta_0) \xrightarrow{\mathscr{D}} N(0, V),$$

where

$$V = \Sigma_1 A^{\tau} B^{-1} \Sigma B^{-1} A \Sigma_1,$$
$$\Sigma_1(\beta_0) = \left\{ A(\beta_0)^{\tau} B(\beta_0)^{-1} A(\beta_0) \right\}^{-1},$$
$$A(\beta_0) = \left. \frac{\partial E \psi(T, Z, \beta)}{\partial \beta} \right|_{\beta_0},$$
$$B(\beta_0) = E \left. \frac{\psi(T, Z, \beta)\psi(T, Z, \beta)^{\tau}}{K(T)} \right|_{\beta_0},$$

and Σ is given in Lemma 1.

To use Theorem 1 to construct confidence interval for parameter β, we have to estimate $A(\beta)$ and $B(\beta)$. Based on the results of Lemma 4, $A(\beta)$ and $B(\beta)$ can be estimated consistently by

$$A_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\nabla_{\beta}\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)} \quad \text{and} \quad B_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)}\right\}^{\otimes 2},$$

respectively. Under some mild regular conditions, it can be shown that

$$\mathcal{R}(\beta) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{ni}(\beta)\right]^{\tau}\left[\frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)W_{ni}(\beta)^{\tau}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{ni}(\beta)\right] + o_p(1), \tag{12}$$

where $W_{ni}(\beta) = V_{ni}\phi(X_i, Z_i, \beta)$. It can be readily shown that $\mathcal{R}(\beta)$ converges in distribution to a weighted sum of Chi-square distributions, as stated in the following theorem.

Theorem 2 *Let Assumptions 1–6 in the Appendix be satisfied. Then,*

$$\mathcal{R}(\beta_0) \xrightarrow{\mathscr{D}} \omega_1\chi_{1,1}^2 + \cdots + \omega_q\chi_{1,q}^2, \tag{13}$$

where the weights ω_j, $1 \leq j \leq q$, are the eigenvalues of $B(\beta_0)^{-1}\Sigma(\beta_0)$, and $\chi^2_{1,j}$ for $1 \leq j \leq q$ are independently distributed Chi-square variables with 1 degree of freedom.

Remark 1 We can give a modification of $\mathcal{R}(\beta_0)$, let

$$\eta_1(\beta_0) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right]^{\tau}\left[\widehat{\Sigma}(\beta_0)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right]$$

$$\eta_2(\beta_0) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right]^{\tau}\left[\widehat{B}(\beta_0)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(\beta_0)\right]$$

and let

$$\widehat{\mathcal{R}}(\beta_0) = \widehat{\xi}(\beta_0)\mathcal{R}(\beta_0),$$

where $\widehat{\xi}(\beta_0) = \eta_1(\beta_0)/\eta_2(\beta_0)$ and $\widehat{B}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta_0)W_{ni}(\beta_0)^{\tau}$. It can be shown $\widehat{\mathcal{R}}(\beta_0)$ has the limiting Chi-square distribution with q degree of freedom.

Corollary 1 *Let $\beta^{\tau} = (\beta_1^{\tau}, \beta_2^{\tau})$, where β_1 are $q_1 \times 1$ vector and β_2 are $q_2 \times 1$ vectors. For $H_0 : \beta_1 = \beta_{1,0}$, the profile empirical likelihood test statistic is*

$$\mathcal{R}_2 = \mathcal{R}(\beta_{1,0}, \tilde{\beta}_{2,0}) - \mathcal{R}(\widehat{\beta}_1, \widehat{\beta}_2),$$

where $\tilde{\beta}_{2,0}$ minimizes $\mathcal{R}(\beta_{1,0}, \beta_2)$ with respect to β_2, $\beta_{1,0}$ is the true value of β_1, and $\widehat{\beta}_e = (\widehat{\beta}_1, \widehat{\beta}_2)$.
Under H_0,

$$\mathcal{R}_2 \to \rho_1\chi^2_{1,1} + \cdots + \rho_{q_1}\chi^2_{1,q_1},$$

where the weights ρ_j, $1 \leq j \leq q_1$, are the eigenvalues of $B_2(\beta_{1,0})\Sigma(\beta_{1,0})$, $\chi^2_{1,j}$ for $1 \leq j \leq q_1$ are independently distributed Chi-square variables with 1 degree of freedom, and B_2 is a positive definite matrix given in the Appendix.

4.2 Generalized method of moments

The GMM approach chooses parameter values such that

$$M_w = \left[\frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta)\right]^{\tau} W \left[\frac{1}{n}\sum_{i=1}^{n} W_{ni}(\beta)\right] \tag{14}$$

is minimized for some positive semi-definite symmetric weight matrix W. In practice, the unknown W is typically replaced by a consistent estimator $\widehat{W}$. The resultant GMM estimator is then

$$\widehat{\beta}_g = \text{argmin}_{\beta} M_{\widehat{w}}(\beta).$$

Theorem 3 *Let Assumptions* 1–6 *in the Appendix be satisfied. Then,*

$$\sqrt{n}(\widehat{\beta}_g - \beta_0) \xrightarrow{\mathscr{D}} N(0, \Sigma_g), \tag{15}$$

where $\Sigma_g = \Sigma_2 A^\tau W \Sigma W A \Sigma_2$ *and* $\Sigma_2 = \{A(\beta_0)^\tau W A(\beta_0)\}^{-1}$.

The choice of W that leads to the most asymptotically efficient GMM estimator is the asymptotic covariance Σ^{-1} defined in Theorem 1, which also results in a "sandwich" covariance Σ_2. If we set $W = B^{-1}$, which is defined in Sect. 4.1, it can be shown that the asymptotic covariance of the resultant GMM estimator coincides with the asymptotic covariance of the EL estimator.

5 Numerical studies

5.1 Simulations

In this section, we carry out simulation studies to evaluate the finite-sample performance of the GMM and EL procedures developed in this paper. The data are generated from the following censored linear regression model, which is similar to the model given in Zhou et al. (2011).

$$T = \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon,$$

where $\beta_1 = 1$, $\beta_2 = 1$, $Z_1 \sim$ Bernoulli distribution with success probability 0.5, $Z_2 \sim U[1, 3]$, ε is generated from the symmetric distribution $\sqrt{2}/4N(0, 1) + t_3/2$ and the censored variable $C \sim U[0, 8.3]$ (for heavy censoring) and $U[0, 25]$ (for light censoring), where C is independent of Z_1, Z_2 and ε. $X = \min(T, C)$, $\Delta = I(T \leq C)$, the corresponding unbiased estimating functions are

$$\psi(X, Z, \beta) = \begin{pmatrix} \psi_1(X, Z, \beta) \\ \psi_2(X, Z, \beta) \\ \psi_3(X, Z, \beta) \\ \psi_4(X, Z, \beta) \end{pmatrix} = \frac{\Delta}{K(X)} \begin{pmatrix} Z_1(X - Z_1\beta_1 - Z_2\beta_2) \\ Z_2(X - Z_1\beta_1 - Z_2\beta_2) \\ Z_1[1/2 - I(X - Z_1\beta_1 - Z_2\beta_2 \leq 0)] \\ Z_2[1/2 - I(X - Z_1\beta_1 - Z_2\beta_2 \leq 0)] \end{pmatrix} \tag{16}$$

with $E\psi_1(\cdot) = 0$ and $E\psi_2(\cdot) = 0$ representing conditions from least squares, and $E\psi_3(\cdot) = 0$ and $E\psi_4(\cdot) = 0$ arising from the median regression. We use the second-order kernel, Gaussian kernel, $l(u) = \exp(-u^2/2)/(2\pi)^{1/2}$ to smooth $\psi_3(\cdot)$ and $\psi_4(\cdot)$. Four estimators are examined, specifically, Koul et al. estimator (1981), least-squares (LS), GMM and EL estimators.

The estimator proposed by Koul et al. (1981) is

$$\widehat{\beta}_K = \left(\sum_{i=1}^{n} Z_i Z_i^\tau \right)^{-1} \sum_{i=1}^{n} Z_i X_{i\widehat{K}}, \tag{17}$$

where $X_{i\widehat{K}} = \frac{\Delta_i X_i}{1-\widehat{K}(X_i)}$, $i = 1, \ldots, n$, and $1 - \widehat{K}(\cdot)$ is the Kaplan–Meier estimator of the censoring distribution. It generalizes ordinary least-squares estimator to censored linear regression model. The LS estimator uses information only from $\psi_1(\cdot) = 0$ and $\psi_2(\cdot) = 0$, while the GMM and EL estimators use information from all the four EEs. Since we use the second-order kernel $l(u)$, Assumption 1 implies $b = o(n^{-1/4})$. Such a bandwidth is of smaller order of magnitude than $o(n^{-1/5})$ which is usually appropriate for minimizing error of curve estimator. Chen and Hall (1993) suggested choices of b in the range between $n^{-1/2}$ and $n^{-3/4}$ which generally provides quite good coverage accuracy. In our simulation study, we selected b through a rule of thumb proposed by Cui et al. (2002) (see also Fan and Yao 2003; Sepanski et al. 1994; Zhou et al. 2008; Zhou and Liang 2009, etc.) and suggested to set $b = c \times \sigma\{\frac{\Delta}{\widehat{K}(X)}(X - Z^\tau \beta)\}n^{-1/3}$, where $\sigma\{X\}$ is the standard variance of X, and c is a suitable constant. We replace β with its LS estimator and set c to $1.5, 2, 2.5, 3, 3.5$. Tables 1 and 2 report the simulation results with light censoring (about 10 %) and heavy censoring (about 30 %), respectively. Each experiment is based on 1000 replicated samples with sample size $n = 200$. The comparisons are in terms of the magnitude of bias in the estimators (BIAS), standard error of the estimators (SE), standard deviation (SD), coverage probability (COV) at the nominal confident level 95 % and the length of confidence interval (LEN) with the same confident level. The coverage probabilities of Koul et al., LS and GMM estimators were constructed using asymptotic normal distribution, while the coverage probability of EL estimator was constructed by the empirical likelihood method.

From Tables 1 and 2, it can be seen that all the four estimators have very small biases, which implies they are asymptotic unbiased and consistent. Meanwhile, the SD (standard deviation) approximates SE (standard error) of the estimator well and coverage probability is close to the nominal confidence level 95 %. The choices of bandwidth have little influence on the results, and the proposed GMM and EL estimators perform better than Koul et al. and LS estimators with smaller SE, SD and shorter LEN, since the prior make use of more information.

In addition, comparing GMM with EL method, it seems GMM estimator has generally less SE and SD, while EL estimator has shorter LEN, especially when censoring rate increases. Besides, the results of Table 2 with heavy censoring rate are very similar as those in Table 1, which implies that the EL and GMM estimators still perform well although the censoring is heavier. The most interest of this paper is reducing SD and SE by proposed censored GMM and EL method, which indeed illustrated by Tables 1 and 2.

Finally, we compare the performance of the proposed GMM and EL estimators with Gehan and Logrank type of rank regression estimators and Buckley–James (B–J) type estimator, respectively. We do not need smoothed technique to use the Gehan estimator, Logrank estimator and B–J estimator. Results are shown in Table 3 below. Zhou (2005) mainly derived a test and a confidence interval based on the rank estimator (Gehan and Logrank type estimators) of regression coefficient in the accelerated failure model. Compared with proposed GMM and EL method from Tables 1, 2 and 3, we can notice that bias of Gehan and Logrank estimators is obviously bigger, especially when censoring rate increases. Again we can find that GMM and EL methods generally perform well with less SD, SE and shorter confidence interval than Gehan, Logrank

Table 1 Simulation results with 10 % censoring

	c	Koul et al.		LS		GMM		EL	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
BIAS	1.5	0.003	−0.005	0.002	−0.003	0.004	−0.004	0.003	−0.003
SE		0.177	0.054	0.128	0.044	0.112	0.038	0.110	0.038
SD		0.183	0.054	0.128	0.043	0.107	0.036	0.108	0.036
COV(%)		95.80	95.30	95.20	95.10	93.10	93.40	94.20	94.30
LEN		0.717	0.212	0.503	0.169	0.421	0.142	0.278	0.121
BIAS	2.0	0.003	−0.005	0.002	−0.003	0.004	−0.003	0.003	−0.003
SE		0.177	0.054	0.128	0.044	0.110	0.038	0.110	0.038
SD		0.183	0.054	0.128	0.043	0.107	0.036	0.108	0.036
COV(%)		95.80	95.30	95.20	95.10	93.60	93.80	94.20	95.10
LEN		0.717	0.212	0.503	0.169	0.420	0.141	0.274	0.114
BIAS	2.5	0.003	−0.005	0.002	−0.003	0.004	−0.003	0.002	−0.002
SE		0.177	0.054	0.128	0.044	0.109	0.038	0.112	0.038
SD		0.183	0.054	0.128	0.043	0.107	0.036	0.109	0.037
COV(%)		95.80	95.30	95.20	95.10	93.40	94.10	94.70	95.50
LEN		0.717	0.212	0.503	0.169	0.419	0.141	0.272	0.109
BIAS	3.0	0.003	−0.005	0.002	−0.003	0.004	−0.003	0.002	−0.002
SE		0.177	0.054	0.128	0.044	0.109	0.037	0.112	0.038
SD		0.183	0.054	0.128	0.043	0.107	0.036	0.109	0.037
COV(%)		95.80	95.30	95.20	95.10	93.20	94.00	94.70	95.80
LEN		0.717	0.212	0.503	0.169	0.419	0.141	0.270	0.105
BIAS	3.5	0.003	−0.005	0.002	−0.003	0.003	−0.003	0.002	−0.002
SE		0.177	0.054	0.128	0.044	0.108	0.037	0.113	0.039
SD		0.183	0.054	0.128	0.043	0.107	0.036	0.108	0.036
COV(%)		95.80	95.30	95.20	95.10	93.50	94.40	94.90	95.70
LEN		0.717	0.212	0.503	0.169	0.419	0.141	0.269	0.104

c is a given constant to choose the bandwidth; "$\widehat{\beta}_1$" and "$\widehat{\beta}_2$" are the estimates of "β_1" and "β_2", respectively; "BIAS", "SE" and "SD" denote the bias, standard error, and standard deviation of the estimator, respectively; "COV" denotes coverage probability at the nominal confident level 95 %, and "LEN" denotes the length of confidence interval with the same confident level; "Koul et al.", "LS", "GMM" and "EL" denote Koul et al., least-squares, GMM and EL estimators, respectively

and Buckley–James type estimators. As we expect that we can improve the efficiency of estimators of the parameters in the AFT model by taking account into auxiliary quantile information, and that implies why the proposed estimators are better than the existed methods.

5.2 A real data example

We illustrate the proposed estimating method with the Stanford Heart Transplant data. These data contain the survival times of 184 heart-transplanted patients with their

Table 2 Simulation results with 30 % censoring

	c	Koul et al.		LS		GMM		EL	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
BIAS	1.5	−0.010	−0.012	−0.006	−0.008	−0.003	−0.008	−0.006	−0.005
SE		0.303	0.081	0.145	0.048	0.127	0.042	0.132	0.043
SD		0.304	0.082	0.143	0.047	0.123	0.041	0.122	0.040
COV(%)		94.90	95.10	94.60	93.60	93.50	93.20	94.20	95.50
LEN		1.190	0.323	0.559	0.185	0.481	0.161	0.321	0.133
BIAS	2.0	−0.010	−0.012	−0.006	−0.008	−0.003	−0.007	−0.008	−0.004
SE		0.303	0.081	0.145	0.048	0.125	0.041	0.132	0.043
SD		0.304	0.082	0.143	0.047	0.123	0.041	0.122	0.041
COV(%)		94.90	95.10	94.60	93.60	94.00	94.10	94.70	95.50
LEN		1.190	0.323	0.559	0.185	0.480	0.160	0.316	0.123
BIAS	2.5	−0.010	−0.012	−0.006	−0.008	−0.003	−0.007	−0.010	−0.004
SE		0.303	0.081	0.145	0.048	0.124	0.041	0.135	0.044
SD		0.304	0.082	0.143	0.047	0.123	0.041	0.123	0.041
COV(%)		94.90	95.10	94.60	93.60	94.90	94.20	95.00	95.40
LEN		1.190	0.323	0.559	0.185	0.481	0.160	0.311	0.125
BIAS	3.0	−0.010	−0.012	−0.006	−0.008	−0.003	−0.006	−0.010	−0.004
SE		0.303	0.081	0.145	0.048	0.123	0.041	0.135	0.044
SD		0.304	0.082	0.143	0.047	0.123	0.041	0.123	0.041
COV(%)		94.90	95.10	94.60	93.60	94.90	94.60	94.50	95.60
LEN		1.190	0.323	0.559	0.185	0.483	0.161	0.307	0.122
BIAS	3.5	−0.010	−0.012	−0.006	−0.008	−0.003	−0.006	−0.010	−0.004
SE		0.303	0.081	0.145	0.048	0.123	0.041	0.136	0.044
SD		0.304	0.082	0.143	0.047	0.124	0.041	0.123	0.041
COV(%)		94.90	95.10	94.60	93.60	94.70	94.30	94.00	95.30
LEN		1.190	0.323	0.559	0.185	0.487	0.162	0.306	0.120

c is a given constant to choose the bandwidth; "$\widehat{\beta}_1$" and "$\widehat{\beta}_2$" are the estimates of "β_1" and "β_2", respectively; "BIAS", "SE" and "SD" denote the bias, standard error, and standard deviation of the estimator, respectively; "COV" denotes coverage probability at the nominal confident level 95 %, and "LEN" denotes the length of confidence interval with the same confident level; "Koul et al.", "LS", "GMM" and "EL" denote Koul et al., least-squares, GMM and EL estimators, respectively

ages at the time of first transplant and their T5 mismatch scores, and details can be seen in Miller and Halpern (1982). Out of these 184 patients, 27 patients did not have T5 scores. And of the remaining 157 patients, the survival times of 55 patients were censored. The cutoff date for the data was in February 1980. It is reasonable to believe that the censoring is dictated by administrative decisions. So, we can estimate the survival function of C by Kaplan–Meier estimator.

But in Miller and Halpern's paper, T5 mismatch score was nonsignificant, so age was only considered in their further analysis. Moreover, 5 of the 157 patients' survival times (T) were less than 10 days, so they were deleted to make $\log_{10} T > 0$. Similarly, in this paper, we use the same dataset with 152 patients, only consider the age covariate,

Table 3 Simulation results for other estimators in AFT model

	Censor (%)	Gahan		Logrank		B–J	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
BIAS	10	0.028	0.030	0.037	0.044	0.001	−0.002
SE		0.114	0.101	0.135	0.115	0.132	0.044
SD		0.115	0.099	0.133	0.115	0.128	0.043
COV(%)		94.60	93.10	93.20	93.60	94.20	95.70
LEN		0.449	0.389	0.521	0.451	0.500	0.170
BIAS	30	0.109	0.114	0.153	0.163	0.001	−0.002
SE		0.136	0.120	0.164	0.138	0.141	0.046
SD		0.136	0.118	0.159	0.138	0.139	0.046
COV(%)		87.50	82.80	84.00	78.10	95.40	96.00
LEN		0.534	0.461	0.623	0.539	0.546	0.182

"$\widehat{\beta}_1$" and "$\widehat{\beta}_2$" are the estimates of "β_1" and "β_2", respectively; "BIAS", "SE" and "SD" denote the bias, standard error, and standard deviation of the estimator, respectively; "COV" denotes coverage probability at the nominal confident level 95 %, and "LEN" denotes the length of confidence interval with the same confident level; "Gehan", "Logrank", and "B–J" denote Gehan, Logrank, and Buckley–James type estimators, respectively

and adopt the same model as in Miller and Halpern (1982), to compare our proposed estimator with theirs. The model is

$$\log_{10} T = \alpha + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \epsilon.$$

Three different c are chosen for the bandwidth parameter, which result in three different smoothing bandwidths. The analysis results can be seen in Table 4.

Buckley–James estimator was given in Miller and Halpern (1982) and Gehan and Logrank type rank regression estimators were obtained by the R-codes given in Zhou (2005). As shown in Table 4, proposed GMM and EL estimators indeed have smaller SD than other estimators. Compared with Koul et al. and LS estimator, GMM and EL estimators look more stable and the estimates are similar to Buckley–James estimator, Gehan estimator and Logrank estimator. Moreover, different choices of smoothing bandwidths have little effect on the results, especially for Age^2 covariate.

6 Discussion

In this article, we proposed a method to estimate the parameters of interest in the AFT model by combining the quantile information with censored least-squares normal equations in the estimating equations. The proposed method is based on the EL and GMM methods, and estimators obtained both have smaller standard error and standard deviation than other estimators such as Koul et al. and LS estimators, which are illustrated in the simulation studies. And their asymptotic properties were studied under some regular conditions. However, there are some problems which need further study. For example, both of the referees ask whether some practical guideline can be used

Table 4 Stanford Heart Transplant data

Estimator	c	Intercept		Age		Age^2	
		$\widehat{\alpha}$	SD($\widehat{\alpha}$)	$\widehat{\beta}_1$	SD($\widehat{\beta}_1$)	$\widehat{\beta}_2$	SD($\widehat{\beta}_2$)
Buckley–James		1.35	0.71	0.107	0.037	−0.0017	0.0005
Gehan		–	–	0.105	0.056	−0.0017	0.0007
Logrank		–	–	0.099	0.048	−0.0016	0.0006
Koul et al.		0.80	1.27	0.039	0.078	−0.0002	0.0011
LS		0.90	0.72	0.098	0.038	−0.0013	0.0005
GMM	0.01	1.37	0.48	0.104	0.025	−0.0017	0.0003
EL		1.73	0.47	0.077	0.024	−0.0014	0.0003
GMM	0.05	1.53	0.46	0.074	0.026	−0.0012	0.0003
EL		1.37	0.46	0.104	0.024	−0.0017	0.0003
GMM	0.10	1.47	0.63	0.069	0.035	−0.0010	0.0005
EL		1.39	0.46	0.101	0.023	−0.0017	0.0003

c is a given constant to choose the bandwidth; "$\widehat{\alpha}$", "$\widehat{\beta}_1$" and "$\widehat{\beta}_2$" are the estimates of intercept term "α" and parameters "β_1" and "β_2", respectively; "SD" denotes the standard deviation of the estimator; "Buckley–James", "Gehan", "Logrank", "Koul et al.", "LS", "GMM" and "EL" denote Buckley–James, Gehan, Logrank, Koul et al., least-squares, GMM and EL estimators, respectively; "–" means no estimates for this term

to choose the bandwidth in the kernel smoothing procedure. There is not a standard method by now as we know, especially for the right censored data. In this paper, the bandwidth is chosen according to the thumb rule, and set $b = c \times \sigma\{\frac{\Delta}{\overline{K(X)}}(X - Z^\tau \beta)\}n^{-1/3}$, where $\sigma\{X\}$ is the standard variance of X. Actually, different c in a large range of possible choices affect little for the results, which can be seen from Tables 1 and 2 of the paper. Besides, the choice of the optimal bandwidth may vary for different datasets, which is a difficult but interesting question, and deserves further study.

Appendix

In this section, we will present the proofs of Theorems 1–3. First, we need some assumptions and symbols. Let $\| \cdot \|$ denote Euclidean norm, $a^{\otimes 2} = aa^\tau$, and $O_p(\cdot)$ denote bound in probability. Assume that $\beta \in \Theta$, where Θ is a tight space.

Assumptions:

1. The selected bandwidth b satisfies the conditions: $b \to 0$, $nb \to \infty$ and $nb^{2r} \to 0$.
2. $L(x)$ is the r-th kernel distribution function such that $\int |x|^r \mathrm{d}L(x) < \infty$.
3. $\tau_s \le \tau_k$, where $\tau_s = \sup\{x : S(x) > 0\}$, $\tau_k = \sup\{x : K(x) > 0\}$, and

$$\int_0^{\tau_s} \frac{\psi(u, z, \beta)^{\otimes 2}}{K(u)} \mathrm{d}F_z(u) < \infty.$$

4. $Q(\beta) = E_z \psi(T, Z, \beta)$ is r-th continuously differentiable in the neighborhood of β_0, the rank of $\partial Q(\beta)/\partial \beta$ is identical to the dimension of parameter β, $\|$

$\partial Q(\beta)/\partial \beta \parallel$ and $\parallel \psi(u, z, \beta) \parallel^3 / K^2(u)$ is bounded by some integrable function $G(u)$ in some neighborhood of β_0.

5. Matrix $B(\beta_0)$ is positive definite.

6. $Q(\beta)$ satisfies the Lipschitz condition in some neighborhood of β_0, that is $\parallel E_z\{\psi(T, Z, \beta) - \psi(T, Z, \beta_0)\}^{\otimes 2} \parallel = O(\parallel \beta - \beta_0 \parallel)$ in some neighborhood of β_0.

Assumptions 1 and 2 are common assumptions used in nonparametric studies, while Assumption 3 is often seen in studies of censored survival data and Assumptions 4–5 are used in empirical likelihood (Qin and Lawless 1994). Assumption 6 can be easily satisfied in many occasions. Note that we only need to smooth the second part of (2). So, in the proof of Lemmas 3, 4, $\phi = \phi_{(2)}$, $\psi = \psi_{(2)}$.

Lemma 1 *Suppose that the Assumption 3 is satisfied. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i)} \psi(X_i, Z_i, \beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma(\beta_0)), \tag{18}$$

where $\Sigma(\beta_0) = (\sigma_{lk}(\beta_0))_{l,k=1,\ldots,q}$ is the covariance matrix with

$$\sigma_{lk}(\beta_0)) = E[\psi_l(T, Z, \beta_0)\psi_k(T, Z, \beta_0)]$$
$$+ E\left[\int_0^{\tau^*} H_l(T, Z, \beta_0, u) H_k(T, Z, \beta_0, u) I(T \geq u) \frac{\lambda^c(u)}{K(u)} \, du\right], \tag{19}$$

where

$$H_k(T, Z, \beta_0, u) = \psi_k(T, Z, \beta_0) - G_k(\beta_0, u),$$
$$G_k(\beta_0, u) = \frac{1}{S(u)} E[\psi_k(T, Z, \beta_0) I(T \geq u)],$$

$\psi_k(T, Z, \beta)$ is the kth element of $\psi(T, Z, \beta)$.

Proof The proof may be constructed along the lines of Bang and Tsiatis (2000). □

Lemma 2 *Suppose that Assumptions 1–3 and 6 are satisfied. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i)} \phi(X_i, Z_i, \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma(\beta_0)),$$

where $\Sigma(\beta_0)$ is given in Lemma 1.

Proof We only need to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i)} \phi(X_i, Z_i, \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i)} \psi(X_i, Z_i, \beta_0) + o_p(1).$$

In fact,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\phi(X_i, Z_i, \beta_0) \Delta_i}{\widehat{K}(X_i)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(X_i, Z_i, \beta_0) \Delta_i}{\widehat{K}(X_i)}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{[\phi(X_i, Z_i, \beta_0) - \psi(X_i, Z_i, \beta_0)] \Delta_i}{\widehat{K}(X_i)}$$

$$\triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi(X_i, Z_i, \beta_0) \Delta_i}{\widehat{K}(X_i)} + J_1.$$

Next, we will proof $J_1 = o_p(1)$. Similar to the argument of Bang and Tsiatis (2000),

$$J_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\phi(T_i, Z_i, \beta_0) - \psi(T_i, Z_i, \beta_0)]$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau^*} [\tilde{\phi}(T_i, Z_i, \beta_0) - \tilde{G}(\beta_0, u)] \frac{d\mathcal{M}_i^c(u)}{K(u)} + o_p(1)$$

$$\triangleq I_1 + I_2,$$

where $\tilde{\phi}(T_i, Z_i, \beta_0) = \phi(X_i, Z_i, \beta_0) - \psi(X_i, Z_i, \beta_0)$, the definition of $\tilde{G}(\beta_0, u)$ similar to the $G_k(\beta_0, u)$ in Theorem 1, $\mathcal{M}_i^c(u)$ is a martingale (More details can be seen in Bang and Tsiatis 2000, p 332). It can be shown that

$$E I_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E [\phi(T_i, Z_i, \beta_0) - \psi(T_i, Z_i, \beta_0)]$$

$$= \sqrt{n} E [\phi(T, Z, \beta_0) - \psi(T, Z, \beta_0)]$$

$$= O_p \left((n b^{2r})^{\frac{1}{2}} \right) = o_p(1). \tag{20}$$

In addition, note that for any constant vector α,

$$\text{Var}[\alpha^\tau I_1] \le E \left\{ \alpha^\tau [\phi(T, Z, \beta_0) - \psi(T, Z, \beta_0)] \right\}^2$$

$$= E \left\{ \int [\alpha^\tau \psi(T, Z^\tau \beta_0 - b\xi) - \alpha^\tau \psi(T, Z, \beta_0)] dL(\xi) \right\}^2$$

$$\le \int E \left\{ \alpha^\tau \psi(T, Z^\tau \beta_0 - b\xi) - \alpha^\tau \psi(T, Z, \beta_0) \right\}^2 dL(\xi).$$

From assumption of $F_z(\cdot)$ and assumption of $L(\cdot)$, we have

$$\text{Var}[\alpha^\tau I_1] = O_p(b). \tag{21}$$

By (20) and (21) we can get $I_1 = o_p(1)$. Now, we consider I_2.

By the property of martingale, we have $EI_2 = 0$ and the kth diagonal element of variance of I_2 is given by

$$E(I_2)^2_{kk} = E \int_0^\tau [\tilde{\phi}_k(T, Z, \beta_0) - \tilde{G}_k(\beta_0, u)]^2 I(T \geq u) \frac{\lambda^c(u)}{K(u)} du$$

$$\leq \int_0^\tau [E\tilde{\phi}_k^2(T, Z, \beta_0) + \tilde{G}_k^2(\beta_0, u)S(u)] \frac{\lambda^c(u)}{K(u)} du.$$

By the results established for I_1, we know that

$$E\tilde{\phi}_k^2(T, Z, \beta_0) = o_p(1),$$

$$\tilde{G}_k^2(\beta_0, u)S(u) = 1/S(u) \left\{ E\tilde{\phi}_k(T, Z, \beta_0) I(T \geq u) \right\}^2 \tag{22}$$

$$\leq 1/S(u)E \left\{ \tilde{\phi}_k(T, Z, \beta_0) I(T \geq u) \right\}^2.$$

Using (22), we have

$$\tilde{G}_k^2(\beta_0, u)S(u) = o_p(1). \tag{23}$$

By (22) and (23) we have that $I_2 = o_p(1)$. Combining this with $I_1 = o_p(1)$, we complete the proof of Lemma 2. □

Lemma 3 *Suppose that Assumptions 1–3 and 6 are satisfied. Then,*

$$E \left\{ \frac{\phi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} = E \left\{ \frac{\psi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} + o(1).$$

Proof

$$E \left\{ \frac{\psi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} = E \frac{\left\{ Z(\frac{1}{2} - I(\log T - Z^\tau \beta \leq 0)) \right\}^{\otimes 2}}{K(T)} = \frac{1}{4} E \frac{Z^{\otimes 2}}{K(T)},$$

$$E \left\{ \frac{\phi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} = E \left\{ \frac{1}{K(T)} \left\{ \int_{-\infty}^{+\infty} \psi(T, Z, \beta - bw) dL(w) \right\}^{\otimes 2} \right\}$$

$$= \int \int E \frac{Z^{\otimes 2}}{K(T)} \left(\frac{1}{2} - I(\log T - Z^\tau \beta + bu \leq 0) \right)$$

$$\times \left(\frac{1}{2} - I(\log T - Z^\tau \beta + bv \leq 0) \right) dL(u) dL(v)$$

$$= \frac{1}{4} E \frac{Z^{\otimes 2}}{K(T)} + o(1).$$

□

Lemma 4 *Suppose that Assumptions* 1–3 *and* 6 *are satisfied. Then,*

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)} \right\}^{\otimes 2} = E\left\{ \frac{\psi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} + o_p(1),$$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\nabla_\beta \phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)} = \nabla_\beta E \psi(T, Z, \beta) + o_p(1). \tag{24}$$

Proof First, we will prove that

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)} \right\}^{\otimes 2} = O_p(1). \tag{25}$$

Note that

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)} \right\}^{\otimes 2} \leq \frac{2}{n} \sum_{i=1}^{n} \left\{ \frac{\phi(X_i, Z_i, \beta)\Delta_i}{K(X_i)} \right\}^{\otimes 2}$$

$$+ \frac{2}{n} \sum_{i=1}^{n} \left\{ \frac{K(X_i) - \widehat{K}(X_i)}{\widehat{K}(X_i)K(X_i)} \right\}^2 \phi(X_i, Z_i, \beta)^{\otimes 2}\Delta_i^2$$

$$\triangleq J_1 + J_2. \tag{26}$$

By the law of large number, we have

$$J_1 = 2E\left\{ \frac{\phi(T, Z, \beta)^{\otimes 2}}{K(T)} \right\} + o_p(1). \tag{27}$$

In addition,

$$J_2 \leq \sup_{0 \leq x \leq X_{(n)}} \left| \frac{K(x) - \widehat{K}(x)}{\widehat{K}(x)} \right|^2 \frac{2}{n} \sum_{i=1}^{n} \left\{ \frac{\phi(X_i, Z_i, \beta)\Delta_i}{K(X_i)} \right\}^{\otimes 2}.$$

Using the fact of Zhou (1991)

$$\sup_{0 \leq x \leq X_{(n)}} \left| \frac{K(x) - \widehat{K}(x)}{\widehat{K}(x)} \right| = O_p(1),$$

we have

$$J_2 = O_p(J_1). \tag{28}$$

By (26), (27), (28) and Lemma 3, we get (25).

Similarly to (27), we can get

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{\widehat{K}(X_i)}\right\}^{\otimes 2} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{K(X_i)}\right\}^{\otimes 2}\\
&+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{K(X_i)-\widehat{K}(X_i)}{\widehat{K}(X_i)K(X_i)}\right\}^{2}\phi(X_i,Z_i,\beta)^{\otimes 2}\Delta_i^{2}\\
&+\frac{2}{n}\sum_{i=1}^{n}\frac{K(X_i)-\widehat{K}(X_i)}{K^2(X_i)\widehat{K}(X_i)}\phi(X_i,Z_i,\beta)^{\otimes 2}\Delta_i^{2}\\
&\triangleq I_1 + I_2 + I_3.
\end{aligned}
\tag{29}
$$

By the law of large number and Lemma 3, we have

$$
I_1 = E\left\{\frac{\psi(T,Z,\beta)^{\otimes 2}}{K(T)}\right\} + o_p(1).
\tag{30}
$$

As for I_2, notice that

$$
I_2 \le \sup_{0\le x\le X_{(n)}}\left|\frac{K(x)-\widehat{K}(x)}{K(x)}\right|^{2}\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{\widehat{K}(X_i)}\right\}^{\otimes 2}.
$$

Using the fact of Gill (1980, p 37)

$$
\sup_{0\le x\le X_{(n)}}\left|\frac{K(x)-\widehat{K}(x)}{K(x)}\right| = o_p(1),
$$

and (25) we get $I_2 = o_p(1)$.

Now, we consider the third part I_3

$$
I_3 \le \sup_{0\le x\le X_{(n)}}\left|\frac{K(x)-\widehat{K}(x)}{K(x)}\right|\frac{2}{n}\sum_{i=1}^{n}\frac{\{\phi(X_i,Z_i,\beta)\Delta_i\}^{\otimes 2}}{\widehat{K}(X_i)K(X_i)},
$$

and

$$
\begin{aligned}
\frac{2}{n}&\sum_{i=1}^{n}\frac{\{\phi(X_i,Z_i,\beta)\Delta_i\}^{\otimes 2}}{\widehat{K}(X_i)K(X_i)}\\
&=\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{K(X_i)}\right\}^{\otimes 2}+\frac{2}{n}\sum_{i=1}^{n}\frac{\{K(X_i)-\widehat{K}(X_i)\}\phi(X_i,Z_i,\beta)^{\otimes 2}\Delta_i^{2}}{\widehat{K}(X_i)K^2(X_i-)}\\
&\le\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{K(X_i)}\right\}^{\otimes 2}+\sup_{0\le x\le X_{(n)}}\left|\frac{K(x)-\widehat{K}(x)}{\widehat{K}(x)}\right|\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\phi(X_i,Z_i,\beta)\Delta_i}{K(X_i)}\right\}^{\otimes 2}.
\end{aligned}
$$

Using (21), similarly to the proof of J_2, we have that $I_3 = o_p(1)$. So, we complete the proof the first result of Lemma 4.

Now, we will consider the second result of Lemma 4, note that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\nabla_\beta\phi(X_i, Z_i, \beta)\frac{\Delta_i}{\widehat{K}(X_i)}\right\} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\nabla_\beta\phi(X_i, Z_i, \beta)\Delta_i}{K(X_i)}\right\}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{K(X_i) - \widehat{K}(X_i)}{\widehat{K}(X_i)K(X_i)}\right\}\nabla_\beta\phi(X_i, Z_i, \beta)\Delta_i \triangleq K_1 + K_2, \quad (31)$$

we have

$$E\left\{\frac{\nabla_\beta\phi(X_i, Z_i, \beta)\Delta_i}{K(X_i)}\right\} = E[\nabla_\beta\phi(T, Z, \beta)]$$
$$= \frac{\partial}{\partial\beta}\int_{-\infty}^{+\infty}\rho(Z^\tau\beta - bw)dL(w) = \nabla_\beta\rho(\beta) + o(1),$$

where $\rho(\beta) \equiv E\psi(T, Z^\tau\beta)$. So, by the law of large number,

$$K_1 = \nabla_\beta\rho(\beta) + o_p(1). \quad (32)$$

Using the fact of Gill (1980, p 37) again

$$|K_2| \leq \sup_{0 \leq x \leq X_{(n)}}\left|\frac{K(x) - \widehat{K}(x)}{K(x)}\right|\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\nabla_\beta\phi(X_i, Z_i, \beta)\Delta_i}{\widehat{K}(X_i)}\right| = o_p(1). \quad (33)$$

Combining (32) and (33), we complete the proof of the second result of Lemma 4. □

Lemma 5 *Suppose that Assumptions 1–3 and 6 are satisfied, then for any β on $\{\beta : ||\beta - \beta_0|| \leq cn^{-\varrho}\}$ where $1/3 < \varrho < 1/2, c$ is some constant, we have*

$$\frac{1}{n}\sum_{i=1}^{n}V_{ni}\phi(X_i, Z_i, \beta) = \frac{1}{n}\sum_{i=1}^{n}V_{ni}\phi(X_i, Z_i, \beta_0) + O_p(n^{-\rho}). \quad (34)$$

Proof We can get the result only by a Taylor expansion. □

Lemma 6 *Suppose that Assumptions 1–6 are satisfied, then $\lambda(\beta) = O_p(n^{-\varrho})$ uniformly on $\{\beta : ||\beta - \beta_0|| \leq cn^{-\varrho}\}$ where $1/3 < \varrho < 1/2, c$ is some constant, and*

$$\lambda(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)W_{ni}(\beta)^\tau\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)\right] + o_p(n^{-\varrho}), \quad (35)$$

uniformly on $\{\beta : ||\beta - \beta_0|| \leq cn^{-\varrho}\}$, where $\lambda(\beta)$ satisfies (10).

Proof By assumptions and the proof of Lemma 3 in Owen (1990), $\max_{1\le i\le n}|V_{ni}^{(0)}\phi$
$(T_i, Z_i, \beta)| = o_p(n^{1/3})$, where $V_{ni}^{(0)} = \Delta_i/K(X_i)$, so we have:

$$\max_{1\le i\le n}|W_{ni}(\beta)| \le \max_{1\le i\le n}|V_{ni}^{(0)}\phi(T_i, Z_i, \beta)| + \max_{1\le i\le n}\left|\frac{\Delta_i(\widehat{K}(T_i) - K(T_i))\phi(T_i, Z_i, \beta)}{\widehat{K}(T_i)K(T_i)}\right|$$

$$\le o_p(n^{1/3}) + \sup_{0\le x\le X_{(n)}}\left|\frac{\widehat{K}(x) - K(x)}{\widehat{K}(x)}\right|\max_{1\le i\le n}|V_{ni}^{(0)}\phi(T_i, Z_i, \beta)|,$$

where $X_{(n)}$ is the largest order statistic. Using the following equality (Zhou 1991),

$$\sup_{0\le x\le X_{(n)}}\left|\frac{K(x) - \widehat{K}(x)}{\widehat{K}(x)}\right| = O_p(1),$$

we have

$$\max_{1\le i\le n}|W_{ni}(\beta)| = o_p(n^{1/3}). \tag{36}$$

Using Lemma 5, similar to the proof of Lemma 3 in Owen (1990), we have $\lambda(\beta) = O_p(n^{-\varrho})$. Using Eq. (10),

$$0 = \frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)\left(1 - Y_i + \frac{Y_i^2}{1 + Y_i}\right),$$

where $Y_i = \lambda(\beta)^\tau W_{ni}(\beta)$, and

$$\max_{1\le i\le n}|Y_i| \le \|\lambda(\beta)\|\max_{1\le i\le n}|W_{ni}(\beta)| = o_p(1).$$

So we have

$$\lambda(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)W_{ni}(\beta)^\tau\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}W_{ni}(\beta)\right] + \beta_n,$$

where

$$\|\beta_n\| \le \frac{1}{n}\sum_{i=1}^{n}\|W_{ni}(\beta)\|^3\|\lambda(\beta)\|^2|1 + Y_i|^{-1} = o_p(n^{-\varrho}).$$

Thus, we complete the proof of Lemma 6. □

Lemma 7 *Suppose that Assumptions 1–6 are hold, then, as $n \to \infty$, with probability 1 $\mathcal{R}(\beta)$ attains its minimum value at some point $\widehat{\beta}_e$ in the interior of the ball $\|\beta - \beta_0\| \le cn^{-\varrho}$, with $\widehat{\beta}_e$ and $\widehat{\lambda} = \lambda(\widehat{\beta}_e)$ satisfying*

$$Q_{1n}(\widehat{\beta}_e, \widehat{\lambda}) = 0, \quad Q_{2n}(\widehat{\beta}_e, \widehat{\lambda}) = 0,$$

where

$$Q_{1n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{V_{ni}\phi(T_i, Z_i, \beta)}{1 + \lambda^{\tau} V_{ni}\phi(T_i, Z_i, \beta)},$$

$$Q_{2n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda^{\tau} \nabla_{\beta} V_{ni}\phi(Y_i, X_i, \beta)}{1 + \lambda^{\tau} V_{ni}\phi(T_i, Z_i, \beta)}.$$

Proof Similar to the proof of Lemma 1 in Qin and Lawless (1994). □

Proof of Theorem 1 Given Lemmas 1–7, the proof of Theorem 1 can be constructed along lines of Theorem 1 of Qin and Lawless (1994). Here, we only give a sketch of the proof. It is easy to show that

$$\begin{pmatrix} \tilde{\lambda} \\ \tilde{\beta}_e - \beta_0 \end{pmatrix} = S_n^{-1} \begin{pmatrix} -Q_{1n}(\beta_0, 0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}, \tag{37}$$

where $\delta_n = \|\widehat{\beta}_e - \beta_0\| + \|\widehat{\lambda}\|$ and

$$S_n = \begin{pmatrix} \nabla_{\lambda} Q_{1n} & \nabla_{\beta} Q_{1n} \\ \nabla_{\lambda} Q_{2n} & 0 \end{pmatrix}_{(\beta_0, 0)} \rightarrow \begin{pmatrix} -B & A \\ A^{\tau} & 0 \end{pmatrix}.$$

From Lemma 2, we have $Q_{1n}(\beta_0, 0) = \frac{1}{n} \sum_{i=1}^{n} V_{ni}\phi(T_i, Z_i, \beta_0) = O_p(n^{-1/2})$. So, we know that $\delta_n = O_p(n^{-1/2})$. Easily we have

$$\sqrt{n}(\widehat{\beta}_e - \beta_0) = S_{22.1}^{-1} A^{\tau} B^{-1} \sqrt{n} Q_{1n}(\theta_0, 0) + o_p(1) \xrightarrow{\mathscr{D}} N(0, V). \tag{38}$$

□

Proof of Theorem 2 The log empirical likelihood ratio multiplied by -2 is given by

$$\begin{aligned} \mathcal{R}(\beta) &= 2 \sum_{i=1}^{n} \log\{1 + \lambda(\beta)^{\tau} W_{ni}(\beta)\} \\ &= 2 \sum_{i=1}^{n} \left\{ \lambda(\beta)^{\tau} W_{ni}(\beta) - \frac{1}{2}(\lambda(\beta)^{\tau} W_{ni}(\beta))^2 \right\} + \gamma_n \\ &= \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta) \right]^{\tau} \left[\frac{1}{n} \sum_{i=1}^{n} W_{ni}(\beta) W_{ni}(\beta)^{\tau} \right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta) \right] + \gamma_n, \end{aligned} \tag{39}$$

where

$$|\gamma_n| \le \|\lambda(\beta)\|^3 \sum_{i=1}^{n} \|W_{ni}(\beta)\|^3 = o_p(1).$$

From Lemma 2, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma(\beta_0)). \tag{40}$$

In addition, by the virtue of Lemma 4,

$$\frac{1}{n} \sum_{i=1}^{n} W_{ni}(\beta_0) W_{ni}(\beta_0)^{\tau} \xrightarrow{p} B. \tag{41}$$

So using (39), (40), (41), we complete the proof of Theorem 2.   □

Proof of Corollary 1

$$
\begin{aligned}
\mathcal{R}_2 &= \mathcal{R}(\beta_{1,0}, \tilde{\beta}_{2,0}) - \mathcal{R}(\widehat{\beta}_1, \widehat{\beta}_2) \\
&= \left[\Sigma(\beta_0)^{-1/2} \sqrt{n} Q_{1n}(\beta_0, 0) \right]^{\tau} \Sigma(\beta_0)^{1/2} B^{-1/2} \\
&\quad \times \left\{ \left(E \frac{\partial W_{ni}}{\partial \beta} \right) \left[\left(E \frac{\partial W_{ni}}{\partial \beta} \right)^{\tau} (E W_{ni} W_{ni}^{\tau})^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta} \right) \right]^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta} \right)^{\tau} \right. \\
&\quad \left. - \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right) \left[\left(E \frac{\partial W_{ni}}{\partial \beta_2} \right)^{\tau} (E W_{ni} W_{ni}^{\tau})^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right) \right]^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right)^{\tau} \right\} \\
&\quad \times B^{-1/2} \Sigma(\beta_0)^{1/2} \left[\Sigma(\beta_0)^{-1/2} \sqrt{n} Q_{1n}(\beta_0, 0) \right] + o_p(1). \tag{42}
\end{aligned}
$$

Denote

$$
\begin{aligned}
B_2 &= B^{-1/2} \left\{ \left(E \frac{\partial W_{ni}}{\partial \beta} \right) \left[\left(E \frac{\partial W_{ni}}{\partial \beta} \right)^{\tau} (E W_{ni} W_{ni}^{\tau})^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta} \right) \right]^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta} \right)^{\tau} \right. \\
&\quad \left. - \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right) \left[\left(E \frac{\partial W_{ni}}{\partial \beta_2} \right)^{\tau} (E W_{ni} W_{ni}^{\tau})^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right) \right]^{-1} \left(E \frac{\partial W_{ni}}{\partial \beta_2} \right)^{\tau} \right\} B^{-1/2}.
\end{aligned}
$$

Similar as Corollary 5 in Qin and Lawless (1994), B_2 is non-negative definite matrix, and then Corollary 1 can be proved easily by Lemma 3 in Qin and Jing (2001).   □

Proof of Theorem 3 It is easy to show that $\widehat{\beta}_g$ is a consistent estimator of β_0 (see, for example, Newey and McFadden 1994, chapter 36, p 2132). By the assumptions, the first-order condition

$$2 A_n(\widehat{\beta}_g)^{\tau} \widehat{W} \frac{1}{n} \sum_{i=1}^{n} W_{ni}(\widehat{\beta}_g) = 0,$$

is satisfied with probability approaching one. Expanding $W_{ni}(\widehat{\beta}_g)$ around β_0 and multiplying through by $\sqrt{n}$, we have

$$\sqrt{n}(\widehat{\beta} - \beta_0) = -\left[A_n(\widehat{\beta}_g)^\tau \widehat{W} A_n(\bar{\beta})\right]^{-1} A_n(\widehat{\beta}_g)^\tau \widehat{W} \sqrt{n} \frac{1}{n} \sum_{i=1}^n W_{ni}(\beta_0),$$

where $\bar{\beta}$ lies between $\widehat{\beta}_g$ and β_0. By Assumption 4 and Lemma 4,

$$A_n(\widehat{\beta}_g) \to^p A, \quad A_n(\bar{\beta}) \to^p A.$$

Thus, we have, in probability, $\left[A_n(\widehat{\beta}_g)^\tau \widehat{W} A_n(\bar{\beta})\right]^{-1} A_n(\widehat{\beta}_g)^\tau \widehat{W} \to^p (A^\tau W A)^{-1} A^\tau W$. The conclusion then follows by the Slutsky theorem. $\qquad\square$

References

Bang, H., Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika*, *87*, 329–343.

Bang, H., Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics*, *58*, 643–649.

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

Chen, S. X., Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, *21*, 1166–1181.

Cui, H. J., He, X. M., Zhu, L. X. (2002). On regression estimators with de-noised variables. *Statistica Sinica*, *12*, 1191–1205.

Fan, J., Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. New York: Springer.

Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica*, *34*, 124.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Heller, G. (2007). Smoothed rank regression with censored data. *Journal of the American Statistical Association*, *102*, 552–559.

Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663–685.

Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association*, *53*, 457–481.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koul, H., Susarla, V., van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *The Annals of Statistics*, *9*, 1276–1288.

Lai, T. L., Ying, Z. (1991). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *The Annals of Statistics*, *19*, 1370–1402.

Lin, D. Y. (2000). Linear regression analysis of censored medical costs. *Biostatistics*, *1*, 35–47.

Lin, D. Y., Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, *80*, 573–581.

Liu, X., Ishfaq, A. (2011). Distribution estimation with smoothed auxiliary information. *Acta Mathematicae Applicatae Sinica, English Series*, *27*, 167–176.

Liu, X., Liu, P., Zhou, Y. (2011). Distribution estimation with auxiliary information for missing data. *Journal of Statistical Planning and Inference*, *141*, 711–724.

Miller, R., Halpern, J. (1982). Regression with censored data. *Biometrika*, *69*, 521–531.

Newey, W. K., McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111–2245.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.

Qin, G., Jing, B.-Y. (2001). Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics*, *28*, 661–673.

Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, *22*, 300–325.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics*, *18*, 303–328.

Robins, J. M., Rotnitzky, A. (1992). *Recovery of information and adjustment for dependent censoring using surrogate markers. AIDS epidemiology-methodological issues* (pp. 297–331). Boston: Birkhäuser.

Sepanski, J. H., Knickerbocker, R., Carroll, R. J. (1994). A semiparametric correction for attenuation. *Journal of the American Statistical Association*, *89*, 1366–1373.

Song, X., Ma, S., Huang, J., Zhou, X. (2007). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics*, *8*, 197–211.

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, *11*, 1871–1879.

Zhao, H., Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*, *84*, 339–348.

Zhou, M. (1991). Some properties of the Kaplan-Meier estimator for independent nonidentically distributed random variables. *The Annals of Statistics, 19*, 2266–2274.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, *92*, 492–498.

Zhou, W., Jing, B. Y. (2003). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Statistica Sinica*, *13*, 83–95.

Zhou, Y., Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics*, *37*, 427–458.

Zhou, Y., Wan, A. T. K., Wang, X. (2008). Estimating equations inference with missing data. *Journal of the American Statistical Association*, *103*, 1187–1199.

Zhou, Y., Wan, A. T. K., Yuan, Y. (2011). Combining least-squares and quantile regressions. *Journal of Statistical Planning and Inference*, *141*, 3814–3828.