

This document is downloaded from DR-NTU, Nanyang Technological University Library, Singapore.

Title	Generalized M-estimation for the accelerated failure time model
Author(s)	Wang, Siyang; Hu, Tao; Xiang, Liming; Cui, Hengjian
Citation	Wang, S., Hu, T., Xiang, L., & Cui, H. (2015). Generalized M-estimation for the accelerated failure time model. <i>Statistics</i> , 50(1), 114-138.
Date	2015
URL	<a href="http://hdl.handle.net/10220/43609">http://hdl.handle.net/10220/43609</a>
Rights	© 2015 Taylor & Francis. This is the author created version of a work that has been peer reviewed and accepted for publication by <i>Statistics</i> , Taylor & Francis. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [ <a href="http://dx.doi.org/10.1080/02331888.2015.1032970">http://dx.doi.org/10.1080/02331888.2015.1032970</a> ].

# Generalized M-Estimation for the Accelerated Failure Time Model

Siyang Wang<sup>1</sup>, Tao Hu<sup>2</sup>, Liming Xiang<sup>3\*</sup> and Hengjian Cui<sup>2</sup>

<sup>1</sup> *School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, 100081, China*

<sup>2</sup> *School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China*

<sup>3</sup> *School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637371*

**Abstract:** The accelerated failure time (AFT) model is an important regression tool to study the association between failure time and covariates. In this paper, we propose a robust weighted generalized M-estimation for the AFT model with right-censored data by appropriately using the Kaplan-Meier weights in the generalized M-type objective function to estimate the regression coefficients and scale parameter simultaneously. This estimation method is computationally simple and can be implemented with existing software. Asymptotic properties including the root-n consistency and asymptotic normality are established for the resulting estimator under suitable conditions. We further show that the method can be readily extended to handle a class of nonlinear AFT models. **Simulation results demonstrate satisfactory finite-sample performance of the proposed estimator.** The practical utility of the method is illustrated by a real data example.

**Key words and phrases:** Accelerated failure time model; Generalized M-estimator; Influence function; Kaplan-Meier weights; Right censoring; Robustness.

---

\* Corresponding author. Tel.: +65 65137451; Fax: +65 65158213.

*E-mail address:* LMXiang@ntu.edu.sg.

## 1. Introduction

In the regression analysis of survival data, an important alternative to the widely used Cox proportional hazards model is the accelerated failure time (AFT) model, which relates the logarithm or a known transformation of the failure time to its covariates. In some situations, the AFT model could be preferred over the proportional hazards model due to its quite direct physical interpretation (see, e.g. Komarek and Lesaffre, 2008).

Parameters in the AFT model are typically estimated using maximum likelihood. Unfortunately, the maximum likelihood estimator may produce misleading results when outliers are present in data or the underlying model is contaminated. The influence of outliers in turn can be quite severe in the AFT model. There is an extensive literature on robust estimation for the AFT model with right-censored data. For example, Ying (1993) and Jin et al. (2003) developed some rank-based methods, and Ying et al. (1995) and Huang et al. (2007) studied a least absolute deviations (LAD) method in the AFT model. The estimators resulting from these methods are resistant to the presence of outliers in the response variable but relatively sensitive to extreme covariate values (also referred to as leverage points). He and Fung (1999) proposed a robust median-based estimator within the framework of M-estimators. In addition, the M-estimation method has attracted wide attention because of its good robustness properties. Zhou (1992) investigated the M-estimation based on weighted data, requiring the censoring time independent of covariates. Ritov (1990), Lai and Ying (1994), Jin (2007) and Zhou (2010) developed various M-estimators and showed their nice theoretical properties. By imposing quantile conditions on the censoring variable, Portnoy (2003), Peng and Huang (2008) and Wang and Wang (2009) proposed ingenious methods of estimating linear conditional quantile functions from censored survival data and studied asymptotic behaviours of the corresponding estimators.

Although the M-estimation is a popular robust method in the AFT model, it does not take into account the presence of leverage points (Hampel et al., 1986). To overcome this problem, Salibian-Barrera and Yohai (2008) proposed a class of high breakdown point S-estimators by minimizing a robust M-estimator of the residual scale, and further built up MM-estimators based on the M- and S-estimators to achieve robustness against both outliers in the response and leverage points. More recently, their S-estimators have been modified to cope with asymmetric error distributions by Locatelli et al. (2010), using a parametric estimation of the error distribution assumed from a local-scale family of distributions.

In this paper, we aim to develop an alternative robust method for estimation in the AFT model

based on the generalized M (GM) estimation originally proposed by Krasker and Welsch (1982) in linear regression. The GM-estimators are attractive because of their properties (Maronna et al., 2006) that 1) with an appropriate choice of the loss function, the influence function is bounded and the breakdown point is positive, and 2) they are easy to compute and are reasonable choices when the number of covariates is small. The t-type M-estimator has been proposed by He et al. (2000). It is a special case of the GM-estimator with redescending scores that can attain a very high breakdown point. Our estimator in this paper follows the idea of the t-type M-estimator and embeds the Kaplan-Meier weights (Stute, 1995 and 1999) to cope with censored observations. We call this new estimator the KMW-GM estimator.

The Kaplan-Meier weights have been used by Huang et al. (2007) and Zhou (2010) to build a LAD-estimator and a M-estimator, respectively, in the AFT model. Their estimators are not scale equivariant unless the scale is known or a scale estimate is available. As pointed out by Zhou (2010), one issue that merits further research is to introduce a scale parameter in the M-estimation procedure for regression coefficients. The proposed KMW-GM estimator aims to extend their results to a more general setting, where both the regression coefficients and scale parameter can be simultaneously estimated in the linear or nonlinear AFT model with right-censored data. Compared to the LAD estimator and Zhou (2010)'s M-estimator, the KMW-GM estimator provides a desirable balance between robustness and efficiency. Moreover, it can be easily implemented with existing statistical software.

The rest of the paper is structured as follows. In Section 2, we present the model and define the KMW-GM estimator. In Section 3, we study the asymptotic properties and influence function of the proposed estimator. Section 4 extends the results obtained in the AFT model to the nonlinear censored regression model. We examine the finite sample performance of the proposed estimator through simulations in Section 5, and illustrate its practical usefulness via a real data analysis in Section 6, followed by a conclusion in Section 7. Proofs of main results and some lemmas are given in the Appendix.

## 2. Generalized M-estimator for Censored Data

We consider the AFT model

$$T = X^T \beta + \varepsilon = x_1 \beta_1 + \cdots + x_d \beta_d + \varepsilon, \quad (2.1)$$

where  $T$  is the logarithm of the failure time,  $X = (x_1, \cdots, x_d)^T$  is a  $d$ -dimensional covariate vector,  $\beta = (\beta_1, \cdots, \beta_d)^T$  is a regression coefficient vector and  $\varepsilon$  is the error term with an unknown

distribution such that  $E(\varepsilon) = 0$  and  $Var(\varepsilon) = \sigma^2$ . When  $T$  is subject to right censoring, let  $C$  denote the logarithm of the censoring time,  $\delta = 1\{T \leq C\}$  the censoring indicator and  $Y = \min\{T, C\}$  the observed time. We assume the data consist of  $n$  independent and identically distributed random triplets  $(Y_i, \delta_i, X_i)$  for  $i = 1, \dots, n$ .

Denote the Kaplan-Meier estimator of the distribution function  $F$  of  $T$  by  $\hat{F}_n$ , which can be written in the form  $\hat{F}_n(y) = \sum_{i=1}^n W_{ni}1\{Y_i \leq y\}$  (Stute and Wang, 1993), where the Kaplan-Meier weights  $W_{ni}$ 's are the jumps of estimator  $\hat{F}_n$ , given by

$$W_{ni} = \frac{\delta_i}{n[1 - \hat{G}_n(Y_i)]}, \quad (2.2)$$

for  $i = 1, \dots, n$ , and  $\hat{G}_n$  is the Kaplan-Meier estimator of  $G$ , the distribution function of  $C$ . This weighting method has been discussed in Zhou (1992) and Huang et al. (2007). Our interest centers on the robust estimation of the  $(d + 1)$ -dimensional parameter vector  $\theta = (\beta^T, \sigma)^T$ . To this end, we define the KMW-GM estimator  $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\sigma}_n)^T$  as the solution which minimizes the [generalized M-type](#) objective function

$$Q_n(\theta) = \sum_{i=1}^n W_{ni} \{ \rho[\omega_i(Y_i - X_i^T \beta)/\sigma] + \log(\sigma) \}, \quad (2.3)$$

where function  $\rho(x)$  can be chosen as [a function satisfying conditions \(A1\)-\(A3\)](#) given in [Section 3.1](#) or as  $\rho_0 = ((v + 1)/2) \log(1 + x^2/v)$  for fixed  $v$  and  $1 \leq v \leq 5$ . We use  $v = 3$  in the numerical studies in this paper. See [He et al. \(2000 and 2004\)](#) for a more detailed discussion on the choice of  $\rho_0(x)$ .  $\omega_i$ 's are known weights corresponding to  $X_i$  and can be taken as

$$\omega(X_i) = \{1 + (X_i - M_n)^T V_n^{-1} (X_i - M_n)/d\}^{-1/2},$$

or Mallows weights

$$\omega(X_i) = \min \left[ 1, \left\{ \frac{b}{(X_i - M_n)^T V_n^{-1} (X_i - M_n)} \right\}^{\alpha/2} \right]$$

with constant  $\alpha, b > 0$ . Here  $M_n$  and  $V_n$  are location and scatter estimators for  $\{X_i, 1 \leq i \leq n\}$  and can be respectively defined by  $Median_i\{X_i\}$  and  $Median_i\{|X_i - Median_i\{X_i\}|\}$  when  $d = 1$ . When  $d > 1$ ,  $M_n$  and  $V_n$  are computed using the minimum covariance determinant (MCD) method ([Rousseeuw and Driessen, 1999](#)).

Denote  $\psi(\cdot) = \rho'_0(\cdot)$  and  $\chi(x) = x\psi(x) - 1$ . The KMW-GM estimator  $(\hat{\beta}_n, \hat{\sigma}_n)$  is then obtained by solving the following system of nonlinear equations:

$$\begin{cases} \sum_{i=1}^n W_{ni} \psi \left( \frac{\omega(X_i)(Y_i - X_i^T \beta)}{\sigma} \right) X_i \omega(X_i) = 0, \\ \sum_{i=1}^n W_{ni} \chi \left( \frac{\omega(X_i)(Y_i - X_i^T \beta)}{\sigma} \right) = 0. \end{cases} \quad (2.4)$$

The numerical solutions of these nonlinear equations can be found by using techniques, such as those implemented by 'nleqslv' in R, the fsolve command in Matlab, the MODLE procedure in SAS and the library MINPACK in FORTRAN. It is important to have a good starting point in this estimation procedure. Because of robustness to outliers and computational convenience, we use the LAD estimator of  $\beta$  based on the Kaplan-Meier weights as initial value  $\beta^{(0)}$  and  $MAD\{|Y_i - X_i^T \beta^{(0)}|\}/0.6745$  as initial value  $\sigma^{(0)}$ .

**Remark 1.** (i) Note that, with a redescending  $\varphi$  function, the estimating equations (2.4) may have multiple roots. The LAD-estimator for the AFT model has been showed to be root-n consistent and asymptotically normal under appropriate assumptions by Huang et al. (2007). Therefore, the LAD estimator can be close to the true value of  $(\beta, \sigma)$  when the sample size is large. For small sample cases, we can utilize several different initial values following the manner given by He et al. (2000) to handle multiple roots problem. These candidate initial values could be obtained using existing robust estimators for the AFT model, such as the LAD, S estimators and so on. If different initial values produce different solutions for the KMW-GM estimator, we then select the one that gives the smallest value of the objective function. When outliers occur in the covariates, careful attention should be paid to the LAD starting values because the LAD estimator is not robust with respect to leverage points.

(ii) In addition, as pointed out by an anonymous referee, the KM weights used in this proposal may have an impact on the robustness of the estimator in the presence of outliers. Denote  $r_i(\beta) = Y_i - X_i^T \beta$ . Following the idea of the MM estimator in Salibian-Barrera and Yohai (2008), an t-type estimation with appropriate initial value can be applied to resolve this problem. Specifically, it defines the estimator  $(\hat{\beta}_n^T, \hat{\sigma}_n)^T$  as the solution minimizing the objective function

$$U(\beta, \sigma) = \frac{1}{n} \sum_{i=1}^n \left\{ \delta_i \left[ \rho_0 \left( \frac{r_i(\beta)}{\sigma} \right) + \log \sigma \right] + \frac{1 - \delta_i}{1 - \hat{F}_n(r_i(\beta))} \int_{r_i(\beta)}^{\infty} \left[ \rho_0 \left( \frac{r_i(\beta)}{\sigma} \right) + \log \sigma \right] d\hat{F}_n(r_i(\beta)) \right\},$$

where  $\hat{F}_n$  is the (conditional) Kaplan-Meier estimator based on  $r_i(\beta)$ . The S-regression estimate  $\tilde{\beta}_n$  calculated as in (2.13) of Salibian-Barrera and Yohai (2008) can be a **good initial estimate**. Other refined versions of the estimator above can be defined following the approach in Section 2.3 of Salibian-Barrera and Yohai (2008).

### 3. Large Sample Properties and Influence Function of the KMW-GM Estimator

### 3.1. Large sample properties

We now present asymptotic results for the KMW-GM estimator  $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\sigma}_n)^T$ . In our discussion hereinafter, we denote by  $\theta_0 = (\beta_0^T, \sigma_0)^T$  the unknown true value of  $\theta$ ,  $H$  the distribution function of  $Y$  and  $F^0(x, t)$  the joint distribution of  $(X, T)$ . Following Stute and Wang (1993), the empirical joint distribution of  $(X, T)$  can be defined by

$$\hat{F}^0(x, t) = \sum_{i=1}^n W_{ni} 1\{X_i \leq x, Y_i \leq t\}.$$

Let  $\tau_Y$  be the least upper bound for the support of  $Y$ , that is,  $\tau_Y = \inf\{y : H(y) = 1\}$ . Similarly,  $\tau_T$  and  $\tau_C$  can be defined for  $T$  and  $C$ . We write

$$\tilde{F}^0(x, t) = \begin{cases} F^0(x, t), & t < \tau_Y, \\ F^0(x, \tau_Y-) + F^0(x, \tau_Y)1\{\tau_Y \in A\}, & t \geq \tau_Y, \end{cases}$$

where  $A$  denotes the set of atoms of  $H$ .

Our next theoretical analysis imposes the following conditions.

(A1)  $\rho(\cdot)$  is symmetric about 0 and continuously increasing on  $[0, \infty)$ ,  $\rho(0) = 0$  and  $\rho(+\infty) = +\infty$ .

(A2)  $\chi(x)$  is an increasing function with  $-1 = \chi(0) < \lim_{|t| \rightarrow \infty} \chi(t) = \kappa > 0$ .

(A3) (i)  $\psi = \rho'$  has a continuous and bounded second-order derivative  $\psi^{(2)}$  and  $\sup_{t \in R^1} (1 + |t|^i) \|\psi^{(i-1)}(t)\| < \infty$  for  $i = 1, 2$ . (ii)  $\text{sgn}(t)E_\varepsilon[\psi(t + \varepsilon\omega(x)/\sigma)] > 0$  and  $E_\varepsilon\psi^2(\varepsilon\omega(x)/\sigma) < \infty$  for any  $\sigma > 0$ , where  $\omega(x)$  is defined as in Section 2 (below (2.3)). (iii) The error  $\varepsilon$  has a symmetric, unimodal distribution  $f_0(y)$ .

(A4) The covariate vector  $X$  is bounded and the right end point of the support of  $X^T\beta_0$  is strictly less than  $\tau_Y$ ;  $\sup_x \|\omega(x)x\| < \infty$ ;  $\Lambda$  defined in (3.2) exists and is non-singular.

(B1)  $T$  and  $C$  are independent and  $P(T \leq C|T, X) = P(T \leq C|T)$ .

(B2) Let  $\Delta F(x) = F(x) - F(x-)$  denote the probability mass of  $F$  at  $x$ .  $\tau_T \leq \tau_C$ , where equality may hold except when  $G$  is continuous at  $\tau_T$  and  $\Delta F(x) > 0$  or  $\tau_T = \tau_C = \infty$ .

We briefly comment on these conditions. Condition (A1) is to provide unique values for certain functionals of  $\psi$  which appears in (2.4). Conditions (A2)-(A4) guarantee that  $\theta_0$  may be identified from a sample of  $(X, Y)$ . (B1) assumes that  $\delta$  is conditionally independent of the covariate  $X$  given the failure time  $Y$ . It also assumes that  $T$  and  $C$  are independent, the same as the assumption for the Kaplan-Meier estimator. Note that (B1) allows the censoring variable to be dependent on covariates. (B2) ensures that the distribution of  $T$  can be estimated over its support, and it is also required for strong consistency. Under condition (B1), Stute and Wang (1993) showed that for any

integrable function  $\varphi(x, t)$ , one has

$$\int \varphi(x, t) d\hat{F}_n^0(x, t) \longrightarrow \int \varphi(x, t) d\tilde{F}^0(x, t) \quad \text{with probability one.} \quad (3.1)$$

Since values of  $Y$  can be observed in the range of  $(0, \tau_Y)$  only, it is possible to estimate  $\int \varphi(x, t) dF^0(x, t)$  consistently only whenif  $\tau_T = \tau_Y$  or if  $\varphi(x, t)$  is zero for  $t \geq \tau_Y$ . Thus the condition (B2) is required for strong consistency. Note that (B2) implies  $\tau_T = \tau_Y$ , and is the necessary and sufficient condition so that  $F^0$  can be estimated consistently on its entire support. More details of the discussion are given by Proposition 1 in Appendix A.

Under these conditions, we can establish the consistency of the proposed estimator, as shown by the following result.

**Theorem 1** (Consistency). *Suppose that (A1)-(A4) and (B1)-(B2) hold. Then  $\hat{\beta}_n \xrightarrow{P} \beta_0$  and  $\hat{\sigma}_n \xrightarrow{P} \sigma_0$  as  $n \rightarrow \infty$ .*

Set  $\varphi_\theta(x, t) = \rho_0 [\omega(x)(t - x^T \beta) / \sigma] + \log(\sigma)$ . The matrix

$$\Lambda = \left\{ \int \frac{\partial^2 \varphi_\theta(x, t)}{\partial \theta_r \partial \theta_s} \Big|_{\theta = \theta_0} dF^0(x, t) \right\}_{1 \leq r, s \leq d+1} \quad (3.2)$$

becomes part of the limit covariance matrix of  $\hat{\theta}_n$  as usual. By condition (A3), the expectation of  $\Lambda$  exists. Put

$$Q(s) = \int_0^{s-} \frac{dG(y)}{\{1 - H(y)\}\{1 - G(y)\}} = \lim_{0 < \epsilon \rightarrow 0} \int_0^{s-\epsilon} \frac{dG(y)}{\{1 - H(y)\}\{1 - G(y)\}}. \quad (3.3)$$

Let  $\varphi$  be a vector-valued function with the  $r$ th component denoted by  $\varphi_r$ . To show the asymptotic normality of  $\hat{\theta}_n$ , we further impose the following two conditions. For  $1 \leq r \leq d + 1$ ,

(B3)  $\int \{\varphi_r(X, Y) \gamma_0(Y) \delta\}^2 dF^0 < \infty$ ; and

(B4)  $\int |\varphi_r(X, Y)| Q^{1/2}(Y) dF^0 < \infty$ .

Condition (B3) is a weak moment assumption on  $\varphi_r$  as used in Stute (1999), while (B4) is used mainly to control the bias of a Kaplan-Meier integral. The function  $Q$  is related to the variance process of the empirical cumulative hazard function for the censored data. The asymptotic normality of the proposed estimator is stated in the following theorem.

**Theorem 2** (Asymptotic Normality). *Suppose that the conditions of Theorem 1 and (B3)-(B4) hold. We have  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Lambda^{-1} \Pi \Lambda^{-1})$ , where  $\Lambda$  is given in (3.2) and  $\Pi$  is defined by (A.3).*

The proofs of Theorems 1 and 2 are given in Appendix A.



**Remark 2.** Condition (B1) that  $T$  and  $C$  are conditionally independent may be relatively strong for some applications. A slightly weaker condition one may wish to assume is that  $T$  and  $C$  are independent given  $X$  (Lopez, 2011). Inspired by the weight approach in (2.2), a natural choice of  $W_{ni}$  would be  $\delta_i/[n(1 - G(Y_i|X_i))]$ , where  $G(y|x) = P(C \leq y|X = x)$ . In practice,  $G(\cdot|X_i)$  is unknown and have to be estimated. Using the local Kaplan-Meier estimator (Gonzalez-Manteiga and Cadarso-Suarez, 1994),  $G(\cdot|X_i)$  can be estimated nonparametrically by

$$\hat{G}_n(t|x) = 1 - \prod_{j=1}^n \left\{ 1 - \frac{B_{nj}(x)}{\sum_{k=1}^n I(Y_k \geq Y_j) B_{nk}(x)} \right\}^{\eta_j(t)},$$

where  $\eta_j(t) = I(Y_j \leq t, \delta_j = 0)$ ,  $B_{nk}(x) = K(\frac{x-x_k}{h_n})/\sum_{i=1}^n K(\frac{x-x_i}{h_n})$ ,  $K$  is a density kernel function on  $R^d$ , and  $h_n \in R^+$  is the bandwidth converging to zero as  $n \rightarrow \infty$ . Under some regularity conditions, asymptotic results similar to Theorems 1-2 can be established using the uniform strong law of large numbers together with an i.i.d. representation of the local Kaplan-Meier estimate (Liang et al., 2012). Note that the [estimation provided in Section 2 can also be applied in this situation. See Appendix B for more detailed justification.](#)

### 3.2. Influence function

The sensitivity of an estimator to small amounts of contamination in the distribution can be measured by its influence function. By definition, the influence function (IF) of a functional estimator  $\theta$  at the model distribution  $F$  is

$$IF(x, t; \theta, F) = \lim_{\epsilon \downarrow 0} \frac{\theta((1 - \epsilon)F + \epsilon\Delta_{(x,t)}) - \theta(F)}{\epsilon},$$

where  $\Delta_{(x,t)}$  is a Dirac measure putting all its mass at  $(x, t)$ . In other words, the IF describes the effect of an infinitesimal contamination at  $(x, t)$  on  $\theta$ , standardized by the mass of the contamination. The estimator  $\theta$  is said to be robust if its IF is bounded (Huber, 1981; Hampel et al., 1986).

According to (2.3), the proposed KMW-GM estimator  $\hat{\theta}_n$  can be also defined as the solution of equations

$$\begin{cases} \sum_{i=1}^n W_{ni} \psi \left( \frac{\omega(X_i)(Y_i - X_i^T \beta(\hat{F}_n^0))}{\sigma(\hat{F}_n^0)} \right) X_i \omega(X_i) = \int \psi \left( \frac{\omega(x)(t - x^T \beta(\hat{F}_n^0))}{\sigma(\hat{F}_n^0)} \right) x \omega(x) d\hat{F}_n^0(x, t) = 0, \\ \sum_{i=1}^n W_{ni} \chi \left( \frac{\omega(X_i)(Y_i - X_i^T \beta(\hat{F}_n^0))}{\sigma(\hat{F}_n^0)} \right) = \int \chi \left( \frac{\omega(x)(t - x^T \beta(\hat{F}_n^0))}{\sigma(\hat{F}_n^0)} \right) d\hat{F}_n^0(x, t) = 0, \end{cases} \quad (3.4)$$

which have the asymptotic functional form

$$\begin{cases} \int \psi \left( \frac{\omega(x)(t - x^T \beta(F^0))}{\sigma(F^0)} \right) x \omega(x) dF^0(x, t) = 0, \\ \int \chi \left( \frac{\omega(x)(t - x^T \beta(F^0))}{\sigma(F^0)} \right) dF^0(x, t) = 0. \end{cases} \quad (3.5)$$

Since all the estimators discussed in this paper are Fisher consistent (see Lemma 2 in Appendix A), notations  $(\beta(F^0), \sigma(F^0))$  and  $(\beta_0, \sigma_0)$  will be used interchangeably as is common practice in the literature. Let  $F_{(x,t),\epsilon}^0 = (1 - \epsilon)F^0 + \epsilon\Delta_{(x,t)}$ . The IFs of  $\beta$  and  $\sigma$  can be consequently found by substituting  $F^0$  with  $F_{(x,t),\epsilon}^0$  in (3.5) and taking derivative with respect to  $\epsilon$  at  $\epsilon = 0$ . That is,  $IF(x, t; F^0, \beta(F^0))$  and  $IF(x, t; F^0, \sigma(F^0))$  satisfy the following system of equations

$$\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} IF(x, t; F^0, \beta(F^0)) \\ IF(x, t; F^0, \sigma(F^0)) \end{pmatrix} = \begin{pmatrix} \psi \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) x\omega(x) \\ \chi \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) \end{pmatrix} \sigma(F^0),$$

where

$$\begin{aligned} D_{11} &= \int \psi' \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) x x^T \omega^2(x) dF^0(x, t), \\ D_{12} &= \int \psi' \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) x \frac{t-x^T\beta(F^0)}{\sigma(F^0)} \omega^2(x) dF^0(x, t), \\ D_{21} &= \int \chi' \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) x^T \omega(x) dF^0(x, t), \\ D_{22} &= \int \chi' \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) \frac{t-x^T\beta(F^0)}{\sigma(F^0)} \omega(x) dF^0(x, t). \end{aligned}$$

By the definition of  $\Lambda$  in (3.2), we have that  $\Lambda = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$ . The expressions of both IFs are then obtained by

$$\begin{pmatrix} IF(x, t; F^0, \beta(F^0)) \\ IF(x, t; F^0, \sigma(F^0)) \end{pmatrix} = \Lambda^{-1} \begin{pmatrix} \psi \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) x\omega(x) \\ \chi \left( \frac{\omega(x)(t-x^T\beta(F^0))}{\sigma(F^0)} \right) \end{pmatrix} \sigma(F^0).$$

Note that this is a heuristic derivation of the IF given that the Gateaux derivative with respect to  $\epsilon$  exists. A rigorous proof can be done along the same line of Huber (Section 2.5, 1981). Under conditions (A1)-(A4) and (B1)-(B4), the estimators  $\beta(\hat{F}_n^0)$  and  $\sigma(\hat{F}_n^0)$  obtained from (3.4) are strongly consistent and asymptotically normal with asymptotic variance  $\int IF(x, t; F^0, \beta(F^0))^2 dF^0(x, t)$  and  $\int IF(x, t; F^0, \sigma(F^0))^2 dF^0(x, t)$ , respectively, where the IFs of  $\beta$  and  $\sigma$  are bounded given conditions (A2) and (A4).

Next, we consider the IFs in the presence of censoring. According to (3.4), the proposed KMW-GM estimator  $\hat{\theta}_n$  can also be defined as the solution of equations

$$\int \varphi(x, t) d\hat{F}_n^0(x, t) = 0 \tag{3.6}$$

where  $\varphi = (\varphi_1, \dots, \varphi_{d+1})^T$  and

$$\varphi_r(x, t) = \begin{cases} \psi(\omega(x)(t - x^T \beta_0)/\sigma_0) x_r \omega, & 1 \leq r \leq d; \\ \chi(\omega(x)(t - x^T \beta_0)/\sigma_0), & r = d + 1. \end{cases}$$

Under condition (B1), it was shown by Stute and Wang (1993) that for any integrable function  $\varphi(x, t)$ ,

$$\int \varphi(x, t) d\hat{F}_n^0(x, t) \longrightarrow \int \varphi(x, t) d\tilde{F}^0(x, t) \quad \text{with probability one.}$$

Similar to Stute (1996), we can show that by condition (B1) and the continuity for the moment

$$\int \varphi(x, t) d\tilde{F}^0(x, t) = \int \varphi(x, t) \exp \left\{ \int_0^{t-} \frac{d\tilde{H}^0(v)}{1 - \tilde{H}(v)} \right\} d\tilde{H}^{11}(x, t) = \int \varphi(x, t) \gamma_0(t) d\tilde{H}^{11}(x, t),$$

where  $\tilde{H}^{11}(x, t) = P\{X \leq x, Y \leq t, \delta = 1\}$ . Using notation  $\gamma_1^\varphi$  and  $\gamma_2^\varphi$  defined by (A.4) and (A.5) in Appendix A, we have

$$\begin{aligned} E[\gamma_1^\varphi(Y_i)(1 - \delta_i)] &= E[\gamma_2^\varphi(Y_i)] \\ &= \int \int \frac{1}{1 - H(t)} 1_{\{t < w\}} \varphi(x, w) \gamma_0(w) d\tilde{H}^{11}(x, w) d\tilde{H}^0(t), \end{aligned}$$

thus  $\int [\gamma_1^\varphi(t)(1 - \delta) - \gamma_2^\varphi(t)] dH(t) = 0$ . Define sub-distributions  $\tilde{H}^0(t) = P(Y \leq t, \delta = 0)$ ,  $\tilde{H}^1(t) = P(Y \leq t, \delta = 1)$  and  $H(t) = \tilde{H}^0(t) + \tilde{H}^1(t)$ . Under condition (B2), we have an asymptotic functional form of (3.6),

$$\begin{aligned} 0 &= \int \varphi(x, t) \gamma_0(t) d\tilde{H}^{11}(x, t) + \int [\gamma_1^\varphi(t)(1 - \delta) - \gamma_2^\varphi(t)] dH(t) \\ &= \int [\varphi(x, t) \gamma_0(t) - \gamma_2^\varphi(t)] d\tilde{H}^{11}(x, t) + \int [\gamma_1^\varphi(t) - \gamma_2^\varphi(t)] d\tilde{H}^0(t). \end{aligned} \quad (3.7)$$

Note that  $\delta$  can only be 0 or 1. The two terms in the last equation of (3.7) can not appear simultaneously. Let  $\tilde{H}_{(x,t),\epsilon}^{11} = (1 - \epsilon)\tilde{H}^{11} + \epsilon\Delta_{(x,t)}$ . The IFs of  $\theta$  can be obtained straightforwardly by substituting  $\tilde{H}_{(x,t),\epsilon}^{11}$  for  $\tilde{H}^{11}$  in (3.7), and then taking derivative with respect to  $\epsilon$  at  $\epsilon = 0$ . When  $\delta = 0$ , we consider  $\tilde{H}^0$  similarly.

By (3.6)-(3.7), the IFs for the corresponding M-estimators of  $\theta$  at censored and uncensored observations are given by

$$IF_1(x, t; \tilde{H}^{11}, \tilde{H}^0, \theta(\tilde{H}^{11}, \tilde{H}^0)) = \Lambda^{-1}[\varphi(x, t) \gamma_0(t) - \gamma_2^\varphi(t)] \sigma_0$$

and

$$IF_0(x, t; \tilde{H}^{11}, \tilde{H}^0, \theta(\tilde{H}^{11}, \tilde{H}^0)) = \Lambda^{-1}[\gamma_1^\varphi(t) - \gamma_2^\varphi(t)] \sigma_0,$$

respectively.

#### 4. Nonlinear censored regression

In this section, we consider an extension of the proposed KMW-GM estimation to a class of nonlinear AFT models, which is a useful compromise between linear AFT and nonparametric models and allows flexible modeling of various data structures. A complete review of the nonlinear models can be found in Seber and Wild (1989). Assume that we observe a sample of independent identically distributed random vectors  $(X_i, T_i)$  in  $(p + 1)$ -dimensional Euclidean space,  $1 \leq i \leq n$ , defined on some probability space  $(\Omega, \mathcal{A}, P)$ .

##### 4.1 KMW-GM estimator for the nonlinear censored regression

Let  $m(x) = E(Y_1|X_1 = x)$ . Suppose that the admissible  $m$  is of the form  $m(x) = f(x, \beta)$ , where  $f(x, \beta)$  is a known function,  $\beta$  is a  $d$ -dimensional parameter, and  $d$  and  $p$  may be different. The nonlinear censored regression assumes that

$$T = f(X, \beta) + \varepsilon \quad \text{with } E(\varepsilon|X) = 0.$$

The KMW-GM estimator is then defined by a solution minimizing

$$S_n^f(\beta, \sigma) = \sum_{i=1}^n W_{ni} \{ \rho_0[\omega(X_i)(Y_i - f(X_i, \beta))/\sigma] + \log(\sigma) \}. \quad (4.1)$$

In fact, such a nonlinear optimization problem is often challenging to solve directly. Instead of optimizing the objective function (4.1) directly, similar to the idea used in Section 2, we can transform the nonlinear optimization problem into a problem of solving a system of nonlinear equations. This allows us to obtain the estimator of  $(\beta, \sigma)$  with existing software for solving nonlinear equations.

##### 4.2 Properties of the KMW-GM estimator for the nonlinear censored regression

To prove the consistency and asymptotic normality of  $\hat{\theta}_n$  in the nonlinear censored regression, we shall make use of the following conditions.

(A4)\* The covariate  $X$  is bounded and the right end point of the support of  $f(X, \beta_0)$  is strictly less than  $\tau_Y$ , and  $\Omega^f$  defined below exists and is non-singular.

(C1)  $\theta \in \Theta_1$  and  $\Theta_1$  is compact.

(C2)  $f(x, \beta)$  is continuous in  $\beta$  for each  $x$  and  $f(x, \beta_1) = f(x, \beta_2)$  if and only if  $\beta_1 = \beta_2$ .

(C3)  $f^2(x, \beta) \leq M(x)$  for some integrable function  $M$ .

(C4)  $f(x, \cdot)$  is twice continuously differentiable .

Conditions (C1) and (C2) guarantee that a minimizer  $\hat{\theta}_n$  exists. (C2) ensures that  $\theta_0$  to be identified from a sample of  $(X, Y)$ . If  $f(x, \cdot)$  admits a continuous extension to the compactification

of  $\Theta_1$ , (C1) becomes superfluous (Richardson and Bhattacharyya, 1986). Conditions (C3)-(C4) and the dominated convergence theorem guarantee that all relevant integrals are continuous in  $\theta$ . It is also needed to show that  $S_n^f$  converges with probability one uniformly in  $\Theta_1$ .

**Theorem 3** (Consistency). *Suppose that (A1)-(A3), (A4)\*, (B1)-(B2) and (C1)-(C3) hold. Then  $\hat{\beta}_n \xrightarrow{P} \beta_0$  and  $\hat{\sigma}_n \xrightarrow{P} \sigma_0$  as  $n \rightarrow \infty$ .*

Updating the conditions (B3) and (B4) in Section 3.1 by

$$(B3)^* \int \{\varphi_r^f(X, Y) \gamma_0(Y) \delta\}^2 dF^0 < \infty \text{ and}$$

$$(B4)^* \int |\varphi_r^f(X, Y)| Q^{1/2}(Y) dF^0 < \infty,$$

we have the following theorem for the asymptotic normality of the KMW-GM estimator in the nonlinear censored regression.

**Theorem 4** (Asymptotic Normality). *Under the condition of Theorem 3, (B3)\*-(B4)\* and (C4) hold. We have  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (\Lambda^f)^{-1} \Pi^f (\Lambda^f)^{-1})$ , where  $\Lambda^f$  and  $\Pi^f$  are defined like  $\Lambda$  and  $\Pi$  in Section 3.1.*

We omit the proofs of Theorems 3 and 4 here as they are similar to those of Theorems 1 and 2.

## 5. Simulation Studies

In this section, we evaluate finite sample performance of the proposed KMW-GM estimator in a variety of settings.

**Simulation 1** (Linear censored regression). In this simulation, the proposed KMW-GM estimator is compared with the KMW-LAD estimator in Huang et al. (2007). We consider an AFT model with a single covariate having the following form:

$$t_i = \beta_1 + \beta_2 x_i + \sigma \varepsilon_i, \tag{5.1}$$

where  $(\beta_1, \beta_2) = (0, 1)$  and  $\sigma = 1$ .

First, we examine the performance of the two estimators given that error term  $\varepsilon_i \sim N(0, 1)$ . We generate values of covariate  $x_i$  from the  $U(0, 6)$  distribution, and censoring times  $C_i$  from the uniform distributions  $U(0, 50)$ ,  $U(0, 24)$  and  $U(0, 10)$  providing censoring rates of 0%, 10% and 40%, respectively.  $C_i$ 's are assumed to be independent of the covariate and failure times. The sample size  $n = 100$  and 500. Simulation results based on 1,000 replications are summarized in Table 1 in terms of the mean estimate and square root of the mean square error. We can see that the proposed KMW-GM estimator always gives lower root mean square errors than the KMW-LAD

estimator, and both estimators have similar performance in terms of mean bias.

[Tables 1]

Next, we conduct simulations for different error distributions and covariates. When the error term is assumed to follow the  $t(3)/\sqrt{3}$  distribution, the corresponding results are reported in Table 2, which shows that the KMW-GM estimator still outperforms the KMW-LAD estimator in terms of root mean square error, similar to the results in Table 1. Instead of the symmetric error distributions, we also consider the standardized log-Weibull error distribution. In particular, we assume that  $x_i \sim N(0, 1)$  and error term  $\varepsilon_i$  follows the standard log-Weibull distribution in model (5.1). Censoring variable  $C_i \sim N(\mu, 1)$  and  $\mu = 0.668$  for a censoring rate around 35%. Results in Table 3 show that in comparison with the KMW-LAD estimator, the KMW-GM estimator has smaller root mean square errors across all simulated scenarios again and comparative mean biases in most scenarios.

[Tables 2-3]

Furthermore, to examine the performance of the two methods in the presence of high leverage points or contamination in both the covariate and error distribution, we assume  $\varepsilon_i \sim N(0, 1)$  and covariate  $x_i \sim N(0, 1)$  being independent of  $\varepsilon_i$ . We set  $\beta_1 = 0$ ,  $\beta_2 = 1.5$  and fix sample size  $n = 100$ . The censoring times  $C_i$  are generated from  $N(\mu, 1)$ . Specifically, we consider following two contamination schemes.

(I) Similar to the simulation setting in Salibian-Barrera and Yohai (2008), we contaminate each generated sample with 10% (10 observations) of outliers at  $(x_0, mx_0)$ , where  $x_0 = 10$  for high leverage points and  $m = 2, \dots, 5$ . We select  $\mu = 1$ , yielding a censoring proportion around 32%. Results of root mean square errors for slope are given in Table 4(I), which shows that the proposed KMW-GM method performs satisfactorily in comparison with the other two methods.

(II) We assume that error term  $\varepsilon_i \sim N(0, 1)$  having 10% contamination by  $N(0, 16)$ , while covariate  $x_i$  follows the  $N(0, 1)$  distribution with 10% contamination by  $N(-3, 1/9)$ .  $\mu = 5, 2$  and 1, giving censoring rates of 0%, 10% and 40%, respectively. As is evident in Table 4(II), the proposed KMW-GM method gives smaller root mean square errors than the KMW-LAD for all scenarios.

[Tables 4(I)-(II)]

In summary, our simulation results demonstrate that the proposed KMW-GM estimator for the regression parameter is quite efficient regardless of the underlying distribution of the error term and across various censoring rate specifications. The proposed method also performs satisfactorily in the presence of high leverage points or contamination in both the covariate and error distribution.

To appreciate the properties of the proposed estimator, we present in Figure 1 boxplots of the KMW-GM estimates of  $\beta_2$  with the standard normal,  $t(3)/\sqrt{3}$  and standard log-Weibull errors. It is seen that all the KMW-GM estimates of  $\beta_2$  are close to the true parameter value. The accuracy of the approximation increases as the censoring rate decreases and the sample size increases. We omit the boxplots for  $\beta_1$  as they show similar trends.

[Figure 1]

In addition, it is often difficult to analytically compute the standard error of the proposed estimator using the asymptotic distribution given by Theorem 2 because of the complex form of the asymptotic covariance  $\Pi$  given in Appendix A. One can estimate the standard error using the weighted bootstrap method (Ma and Kosorok, 2005) instead. We perform the following simulation to assess the bootstrap estimate. Consider an AFT model with the same parameter setting as in the previous simulation. The error term is assumed to be normally or t-distributed with mean 0 and standard deviation 1. The standard errors are computed using the weighted bootstrap in comparison with the sample standard deviations.

Results are listed in Table 5 based on 1,000 replicates and 1,000 bootstrap for each sample with a fixed censoring rate of 40%. As expected, the bootstrap standard errors of the KMW-GM estimates are very close to their empirical counterparts.

[Table 5]

**Simulation 2** (Nonlinear censored regression). Here we examine the validity of our results in the nonlinear censored regression model for finite samples. The underlying admissible nonlinear regression function is assumed to be

$$f(x, \beta) = (\beta_1 + \beta_2)^{-1} \exp(\beta_1 x_1 + \beta_2 x_2).$$

Data  $(X_i, T_i)$ ,  $1 \leq i \leq n$ , are generated as follows: each  $X_i = (X_i^1, X_i^2)$  consists of two independent random variables from the uniform distribution on  $[0, 3]$ . The true parameters are set to be

$\theta_0 = (\beta_{01}, \beta_{02}, \sigma_0)^T = (0.5, 0.3, 1)^T$  so that

$$T_i = \frac{5}{4} \exp(0.5X_i^1 + 0.3X_i^2) + \varepsilon_i, \quad 1 \leq i \leq n.$$

We choose sample size  $n = 100$ , and set the censoring distribution  $G$  to be the uniform distribution on  $[0, 60]$ ,  $[0, 20]$  and  $[0, 12]$  corresponding to censoring rate of 0%, 10% and 40%, respectively.  $C_i$  is taken to be independent of  $(X_i, T_i)$ . When the sample size changes, the distribution of  $G$  also changes. The error term  $\varepsilon_i$  is assumed to be independent of  $X_i$  and generated from the  $N(0, 1)$ ,  $t(3)/\sqrt{3}$  and  $t(5)/\sqrt{5/3}$  distributions, respectively, where  $E(\varepsilon_i) = 0$  and  $Var(\varepsilon_i) = 1$ .

For each combination of the simulation parameters, we generate 1000 replicates of data. For the minimization of  $S_n^f$ , we apply a modified Gauss-Newton algorithm, which can be found in the R package. The simulated means and square root of the mean square errors are reported in Tables 6-8 for [differently distributed errors](#). It becomes clear that the bias slightly increases as censoring becomes more and more substantial. The proposed KMW-GM estimation performs satisfactorily for finite samples with different sample sizes, degrees of censoring and error distributions.

[Tables 6-8]

**Simulation 3** (Comparison with censored quantile regression methods). We further conduct simulations to compare the proposed KMW-GM estimation with the existing censored quantile regression methods developed by Portnoy (2003), Peng and Huang (2008) and Wang and Wang (2009). Following the same setup used in Wang and Wang (2009) (example 1), we generate data  $(x_i, T_i, C_i)$  from the model  $T_i = b_0 + b_1x_i + \varepsilon_i$ ,  $i = 1, \dots, n$ , where  $\beta_0 = 3$ ,  $\beta_1 = 5$ ,  $x_i \sim U(0, 1)$ . To see how the robust estimators protect us from gross errors in the data, we assume that  $\varepsilon$  follows one of the following distributions:  $N(0, 1)$ ,  $t(3)/\sqrt{3}$  and  $Laplace(0, \sqrt{0.5})$ . The censoring variable  $C_i \sim U(0, 14)$ , resulting in 40% censoring at the median. The observed response variable is  $Y_i = \min(T_i, C_i)$ .

Table 9 summaries the results for two different sample sizes  $n = 200$  and  $500$  using four different methods: CRQP (Portnoy, 2003,  $\tau = 0.5$ ), CRQPH (Peng and Huang 2008,  $\tau = 0.5$ ), LCRQ (Wang and Wang, 2009,  $\tau = 0.5$ ) and KMW-GM (the proposed method with the local Kaplan-Meier weight). In this simulation, the CRQP and CRQPH were calculated using the function `crq` in R package `quantreg`, and the LCRQ was calculated using the R code available at <http://www4.stat.ncsu.edu/wang/research/software/LCRQ.R>.

Results in Table 9 show that the KMW-GW estimator performs very similar to the other estimators based on censored quantile regression methods when  $\varepsilon$  follows the  $Laplace(0, \sqrt{0.5})$



distribution. When  $\varepsilon$  follows  $N(0, 1)$  or  $t(3)/\sqrt{3}$ , the KMW-GM method still achieves comparable performance to the other three methods and can protect us against errors in data.

[Table 9]

## 6. Application

For illustration, we apply the KMW-GM estimation to the Heart data set (Kalbfleisch and Prentice, 2002) on 155 heart transplant recipients, including their age, mismatch score and their time to death or censoring (survival). The data set is available in the “survival” library of R, and has been analyzed by Salibian-Barrera and Yohai (2008) and Locatelli and Marazzi and Yohai (2011) using model  $y = \alpha + \beta x + \sigma u$ , where  $y = \log(\text{time})$  and  $x = \text{age}$ . In this section, we consider the AFT model  $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \sigma \varepsilon$ , where  $y = \log(\text{time})$ ,  $x_1 = \text{age}$  and  $x_2 = \text{mismatch score}$ . First, we present the standard residuals in Figure 2 computed using the uncensored data only and the complete data set, respectively. Based on Figure 2, the data points with ID 2, 16 and 21 are found to be possible outliers, which are consistent with the results from previous studies.

[Figure 2]

We apply the proposed robust method to estimate the parameters in the model. Resulting estimates obtained with or without those outliers are presented in Table 10. Figure 2 suggests that the residual distribution is quite asymmetric. It can be seen that with the full data set the KMW-GM estimate of  $\beta_1$  is close to zero, which indicates that variable age slightly affects the mean survival time. We observe that both the KMW-GM and KMW-LAD estimates change little after removal of the outliers. To assess such small changes in the estimates, we further calculate the relative variation of the estimates with or without outliers. Suppose that  $\theta_1$  is the estimate based on the full data set, and  $\theta_2$  is the estimate based on the reduced data set (without outliers). We then define

$$\text{relative variation} = \frac{|\theta_1 - \theta_2|}{|\theta_1|}.$$

Results of the relative variation are provided in Table 11. It is seen that the KMW-LAD estimates vary a lot for some particular parameters after removal of the outliers, while the KMW-GM estimates almost keep unchanged for both the regression and scale parameters, thus illustrating the strong robustness of the KMW-GM estimator. We also observe that standard errors of the KMW-GM estimates are consistently less than those of the KMW-LAD estimates, exhibiting the relative efficiency of the KMW-GM estimator proposed.

## 7. Concluding Remarks

In this paper, we propose a robust method to estimate the regression coefficients and the scale parameter simultaneously in the AFT model with right-censored data. We develop the KMW-GM estimation in which the Kaplan-Meier weights are embedded within the generalized M-estimation. Under suitable assumptions, the resulting estimator is consistent and asymptotically normal, and has a bounded influence. Simulation results show that in comparison with the two existing robust approaches, the KMW-GM estimation performs satisfactorily in terms of robustness and efficiency. The real data analysis also demonstrates the effectiveness of the method. According to our numerical experiments, the algorithm is very stable and can yield estimates within reasonable computing time.

The t-type estimator is a special case of the GM estimator with redescending scores that can attain a very high breakdown point. However, the weights are based on a high breakdown point multivariate location-scale estimator (e.g., the minimum volume ellipsoid estimator). For this reason, the t-type estimator loses the computational advantage of a typical GM estimator when the number of covariates is large.

As shown in the proof of Lemma 2 in Appendix A, the symmetric error distribution condition in (A3) is mainly used to obtain the Fisher consistency. In finite sample cases, we have done simulations with this condition relaxed. Our simulation results in Table 3 show that the KMW-GM estimator performs favorably under the standard log-Weibull error distribution. The proposed KMW-GM estimator can be regarded as a reasonable alternative to the existing robust estimators with convenient implementation using existing packages, given that the number of covariates and the contamination fraction are small, and the error distribution is symmetric.

In addition, we assume that the scale parameter  $\sigma$  is a positive constant for technical consideration. This rules out the possibility of conditional heteroskedasticity, which is allowed in some works in the literature (see, e.g. Ying et al., 1995; Honore et al., 2002). We also require a mean 0 condition for the error term in assumption (A3), which can be somehow circumvented by imposing a median condition or quantile condition (see, e.g. Ying et al. 1995; Honore et al. 2002; Portnoy, 2003; Peng and Huang, 2008; Wang and Wang, 2009). How to weaken these assumptions would be another interesting topic for our future research.

It is noted that the proposed KMW-GM estimation requires independence between censoring

time and event time. This assumption is generally reasonable in many survival studies. When the independence assumption is clearly violated, the proposed method should be applied with more caution.

## Acknowledgement

We are grateful to Editor Prof. Roland Fried, an Associate Editor and two anonymous referees for their very constructive comments and suggestions which helped improve the paper greatly. Wang's research is supported by a grant for Young PhD Researchers from CUFE (QBJ1423). Hu's research is partially supported by the Natural Science Foundation of China (11201317, 11371062). Xiang's research is partially supported by the Singapore MOE AcRF (RG30/12, ARC29/14).

## Appendix A: Proofs of Theorems 1-2

Similar to Stute and Wang (1993) and Zhou (2010), the following proposition and two lemmas are needed for the proofs of our theoretical results.

**Proposition 1. (Strong law of large numbers)** *Under (B1)-(B2), for any  $\varphi$  such that  $\int |\varphi(x, t)| dF(x, t) < \infty$ , it follows that*

$$\int \varphi(x, t) d\hat{F}_n^0(x, t) \longrightarrow \int \varphi(x, t) dF^0(x, t) \quad \text{with probability one.}$$

Moreover,

$$\sup_{-\infty < x < \infty, -\infty < t \leq \tau_Y} |\hat{F}_n^0(x, t) - F^0(x, t)| \longrightarrow 0 \quad \text{with probability one.}$$

Technically, the consistency and asymptotic normality of  $\hat{\theta}_n$  will be shown by first considering a linear approximation of  $\hat{\theta}_n$  and then applying the strong law of large numbers and the central limit theorem to Kaplan-Meier integrals  $\int \varphi d\hat{F}_n^0$  with a proper vector-valued function  $\varphi$ .

Consider

$$S_n(\theta) = \int \varphi_\theta(x, t) d\hat{F}_n^0(x, t)$$

with  $\varphi_\theta(x, t) = \rho_0 [\omega(x)(t - x^T \beta) / \sigma] + \log(\sigma)$ . Recall  $G$ ,  $H$  and  $\tau_Y$ ,  $1 - H = (1 - F)(1 - G)$  by condition (B1). Clearly, there will be no data beyond  $\tau_Y$ . So, if  $\int \varphi dF^0$  is a parameter of interest, the best we can hope for is to consistently estimate the integral restricted to  $t \leq \tau_Y$ . (3.1) asserts that with probability one,

$$\lim_{n \rightarrow \infty} \int \varphi d\hat{F}_n^0 = \int_{\{t < \tau_Y\}} \varphi(x, t) d\tilde{F}^0(x, t) + 1_{\{\tau_Y \in A\}} \int_{\{t = \tau_Y\}} \varphi(x, \tau_Y) d\tilde{F}^0(x, t),$$

where  $A$  is the set of  $H$  atoms, possibly empty. Regarding  $S_n(\theta)$ , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n(\theta) &= \int_{\{t < \tau_Y\}} [\rho_0(\omega(x)(t - x^T \beta)/\sigma) + \log(\sigma)] d\tilde{F}^0(x, t) \\ &\quad + 1_{\{\tau_Y \in A\}} \int_{\{t = \tau_Y\}} [\rho_0(\omega(x)(\tau_Y - x^T \beta)/\sigma) + \log(\sigma)] d\tilde{F}^0(x, t). \end{aligned} \quad (A.1)$$

For most lifetime distributions considered in the literature,  $\tau_T = \tau_C = \infty$  and therefore  $\tau_Y = \infty$ . In this case, the right-hand side of equation (A.1) is equal to  $E[\rho_0(\omega(X)(T - X^T \beta)/\sigma) + \log \sigma]$ . For other cases that the identifiability function in condition (A3) has to be modified accordingly. For ease of notation, we assume without further mention that the limit is  $E[\rho_0(\omega(X)(T - X^T \beta)/\sigma) + \log \sigma]$ .

**Lemma 1.** *If conditions (A1)-(A3) are satisfied, there exist  $\sigma_1, \sigma_2 > 0$  satisfying  $\sigma_1 \leq \sigma_0 \leq \sigma_2$ .*

**Proof.** This lemma can be proved by using similar proof of Lemma 2 in Cui (2004).

**Lemma 2.** *Under the conditions (A1)-(A4) and (B1)-(B2),*

$$(\beta_0^T, \sigma_0)^T = \arg \min_{\beta, \sigma} E_{F^0} \{ \rho[\omega(X)(T - X^T \beta)/\sigma] + \log(\sigma) \},$$

and  $(\beta_0^T, \sigma_0)^T$  is unique.

**Proof.** Let

$$\begin{aligned} h(\beta, \sigma) &= E_{F^0} \{ \rho[(\varepsilon + X^T \beta_0 - X^T \beta)\omega(X)/\sigma] - \rho(\varepsilon\omega(X)/\sigma) \} + E_{F^0} [\rho(\varepsilon\omega(X)/\sigma) + \log(\sigma)] \\ &:= h_1(\beta, \sigma) + h_2(\beta, \sigma). \end{aligned}$$

By condition (A3), we note that

$$\begin{aligned} h_1(\beta, \sigma) &= E_{F^0} \int_{\varepsilon\omega(X)/\sigma}^{(\varepsilon + X^T \beta_0 - X^T \beta)\omega(X)/\sigma} \psi(t) dt = E_{F^0} \int_0^{X^T(\beta_0 - \beta)\omega(X)/\sigma} \psi(t + \varepsilon\omega(X)/\sigma) dt \\ &= E_X \left\{ \int_0^{x^T(\beta_0 - \beta)\omega(x)/\sigma} E_\varepsilon \psi(t + \varepsilon\omega(x)/\sigma) dt \right\}, \end{aligned}$$

and  $h_1(\beta_0, \sigma) = 0$ . Since  $\varepsilon$  has a symmetric distribution and  $\psi(t) \geq 0$  ( $t \geq 0$ ), we can get  $E_\varepsilon[\psi(t + \varepsilon\omega(x)/\sigma)] = E_\varepsilon[\psi(t + |\varepsilon\omega(x)/\sigma|)] \geq 0$  for  $t \geq 0$ . With odd function  $\psi$ , we then have

$$h_1(\beta, \sigma) = E_X \left\{ \int_0^{|x^T(\beta_0 - \beta)\omega(x)/\sigma|} E_\varepsilon \psi(t + \varepsilon\omega(x)/\sigma) dt \right\}.$$

According to the monotonicity of  $\rho(\cdot)$  on  $[0, \infty)$  and  $\rho(\infty) = \infty$ , we know that  $E_\varepsilon \psi(t + |\varepsilon\omega(x)/\sigma|) > 0, P.$ , for  $t \geq 0$ . Assume that there exists  $\tilde{\beta} \neq \beta_0$  such that  $h_1(\tilde{\beta}, \sigma) = 0$ , then  $E_\varepsilon \psi(t + |\varepsilon\omega(x)/\sigma|) = 0, P.$  when  $t \geq 0$ , which contradicts with the conclusion above. In other words,  $\beta_0$  is unique.

Now we consider the uniqueness of  $\sigma_0$ . Because  $E_\varepsilon[\chi(\varepsilon\omega(x)/\sigma_0)] = 0$ , we just need to verify that  $\sigma_0$  is the unique solution of  $E_\varepsilon[\chi(\varepsilon\omega(x)/\sigma)] = 0$ . Assume there exists  $\tilde{\sigma}$  satisfying  $E_\varepsilon[\chi(\varepsilon\omega(x)/\tilde{\sigma})] = 0$ . Without loss of generality, suppose that  $\sigma_0 < \tilde{\sigma}$ . We have  $E_\varepsilon[\chi(\varepsilon\omega(x)/\sigma_0) - \chi(\varepsilon\omega(x)/\tilde{\sigma})] = 0$ . Let  $f_0(\cdot)$  denote the density function of  $\varepsilon$ . Furthermore,

$$\int_0^\infty [\chi(y\omega(x)/\sigma_0) - \chi(y\omega(x)/\tilde{\sigma})]f_0(y)dy + \int_0^\infty [\chi(-y\omega(x)/\sigma_0) - \chi(-y\omega(x)/\tilde{\sigma})]f_0(y)dy = J_1 + J_2 = 0.$$

$J_1, J_2$  are both nonnegative, thus  $J_1 = J_2 = 0$ . Considering conditions (A2) and (A3), we know that  $\sigma_0 = \tilde{\sigma}$ .  $\square$

### Proof of Theorem 1 (Consistency).

Following from the discussion in Section 3.1, under (B1) and (B2) and by Proposition 1, we have for any measurable function  $\varphi$ ,

$$S_n^\varphi = \sum_{i=1}^n W_{ni}\varphi(X_i, Y_i) \longrightarrow S^\varphi \equiv \int \varphi d\tilde{F}^0, \quad a.s. \quad (A.2)$$

provided that  $\int |\varphi| d\tilde{F}^0$  is finite. Applying this result to  $\varphi_\theta(x, t) = \rho_0[\omega(x)(t - x^T\beta)/\sigma] + \log(\sigma)$ , when  $\tau_T < \tau_C$  or  $\tau_Y = \infty$ , we obtain

$$S_n(\theta) = \sum_{i=1}^n W_{ni}\varphi_\theta(X_i, Y_i) \longrightarrow S(\theta), \quad a.s., \text{ for any fixed } \theta \in R^{d+1},$$

where

$$S(\theta) = E [\rho_0(\omega(X)(T - X^T\beta)/\sigma) + \log(\sigma)].$$

Note that  $\varphi_\theta(x, t)$  is a convexity function of  $\theta = (\beta^T, \sigma)^T$ . By the convexity lemma of Pollard (1991), for any compact set  $K$  in a convex open subset of  $R^{d+1}$ ,

$$\sup_{\theta \in K} |S_n(\theta) - S(\theta)| \xrightarrow{P} 0 \quad (A.3)$$

Under the conditions (A1)-(A4) and (B1)-(B2) and by lemmas 1 and 2, we get  $\hat{\beta}_n \xrightarrow{P} \beta_0$  and  $\hat{\sigma}_n \xrightarrow{P} \sigma_0$ . This completes the proof of Theorem 1.  $\square$

To show the result of Theorem 2, we need further notation and definitions. Define sub-distribution functions

$$\tilde{H}^{11}(x, t) = P\{X \leq x, Y \leq t, \delta = 1\} \quad \text{and} \quad \tilde{H}^0(t) = P(Y \leq t, \delta = 0).$$

These functions may be consistently estimated by their empirical counterparts. Let

$$\gamma_0(t) = \exp \left\{ \int_0^{t^-} \frac{d\tilde{H}^0(y)}{1 - H(y)} \right\} = \lim_{0 < \epsilon \rightarrow 0} \exp \left\{ \int_0^{t-\epsilon} \frac{d\tilde{H}^0(y)}{1 - H(y)} \right\}$$

and for every real-valued function  $\varphi$ ,

$$\gamma_1^\varphi(t) = \frac{1}{1-H(t)} \int 1_{\{t < w\}} \varphi(x, w) \gamma_0(w) d\tilde{H}^{11}(x, w), \quad (\text{A.4})$$

$$\gamma_2^\varphi(t) = \int \int \frac{1_{\{v < t, v < w\}} \varphi(x, w) \gamma_0(w)}{[1-H(v)]^2} d\tilde{H}^0(v) d\tilde{H}^{11}(x, w). \quad (\text{A.5})$$

According to Stute and Wang (1993),  $\gamma_0 = (1-G)^{-1}$  for continuous  $H$ . Stute (1995) showed that under weak moment assumptions on  $\varphi$ , in probability,

$$\begin{aligned} \int \varphi d(\hat{F}_n^0 - F^0) &= n^{-1} \sum_{i=1}^n \{ \varphi(X_i, Y_i) \gamma_0(Y_i) \delta_i + \gamma_1^\varphi(Y_i) (1 - \delta_i) - \gamma_2^\varphi(Y_i) \} \\ &\quad + o(n^{-1/2}) \equiv n^{-1} \sum_{i=1}^n \xi_i^\varphi + o(n^{-1/2}), \end{aligned} \quad (\text{A.6})$$

where  $\xi_i$ 's are independently and identically distributed with mean zero. [In this paper we shall have to consider  \$\varphi\_r\$ ,  \$1 \leq r \leq d+1\$  at the same time](#), namely

$$\varphi_r(x, t) = \begin{cases} \psi(\omega(x)(t - x^T \beta_0)/\sigma_0) x_r \omega, & 1 \leq r \leq d; \\ \chi(\omega(x)(t - x^T \beta_0)/\sigma_0), & r = d+1. \end{cases}$$

Set

$$\Pi = (\sigma_{rs})_{1 \leq r, s \leq d+1} \text{ with } \sigma_{rs} = \text{cov}(\xi_1^{\varphi_r}, \xi_1^{\varphi_s}). \quad (\text{A.7})$$

If  $\gamma_0$ ,  $\gamma_1^{\varphi_r}$  and  $\gamma_2^{\varphi_r}$  were known, then we could estimate  $\Pi$  using the sample covariance  $\Pi_n$  of  $\xi$ 's. However,  $\gamma_0$ ,  $\gamma_1^{\varphi_r}$  and  $\gamma_2^{\varphi_r}$  are often unknown in practice. Note that they are functions of  $H$  and can be estimated by their empirical counterparts.

**Proof of Theorem 2 (Asymptotic Normality).** Let  $\theta = (\beta^T, \sigma)^T \in \Theta_1$ . There exists some  $\theta'_n$  between  $\hat{\theta}_n$  and  $\theta_0$  such that

$$n^{1/2}(\hat{\theta}_n - \theta_0) = -A_n^{-1} n^{1/2} \frac{\partial S_n(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \quad (\text{A.8})$$

with

$$A_n = \frac{\partial^2 S_n(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta'_n}.$$

Put

$$b_n = n^{1/2} \frac{\partial S_n(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} := ((b_n^1)^T, b_n^2)^T.$$

We show the asymptotic normality of  $b_n$  first. Note that

$$\begin{aligned}
b_n^1 &= -n^{1/2} \sum_{i=1}^n W_{ni} \psi [\omega(X_i)(Y_i - X_i^T \beta_0)/\sigma_0] X_i \omega(X_i), \\
b_n^2 &= -n^{1/2} \sum_{i=1}^n W_{ni} \chi [\omega(X_i)(Y_i - X_i^T \beta_0)/\sigma_0].
\end{aligned}$$

Denote

$$\varphi(x, t) = [\psi(\omega(x)(t - x^T \beta_0)/\sigma_0) x^T \omega(x), \chi(\omega(x)(t - x^T \beta_0)/\sigma_0)]^T.$$

Rewrite

$$b_n = -n^{1/2} \int \varphi(x, t) \hat{F}_n^0(dx, dt).$$

If  $\varphi$  is a vector-valued function, we write  $\varphi = (\varphi_1, \dots, \varphi_{d+1})$ . From Theorem 1.1 of Stute (1996), we obtain that in probability

$$\begin{aligned}
\int \varphi_r(x, t) \hat{F}_n^0(dx, dt) &= n^{-1} \sum_{i=1}^n \varphi_r(X_i, Y_i) \gamma_0(Y_i) \delta_i \\
&\quad + n^{-1} \sum_{i=1}^n \{\gamma_1^{\varphi_r}(Y_i)(1 - \delta_i) - \gamma_2^{\varphi_r}(Y_i)\} + o(n^{-1/2}), \quad 1 \leq r \leq d+1
\end{aligned}$$

where  $\gamma_1^{\varphi_r}$  and  $\gamma_2^{\varphi_r}$  are given in (A.4) and (A.5), respectively. The second sum consists of independent and identically distributed random variables with zero mean. As to the first sum, note that the definition of the GM estimator,

$$\begin{aligned}
E\{\varphi_r(X_1, Y_1) \gamma_0(Y_1) \delta_1\} &= E[\psi(\omega(X_1)(Y_1 - X_1^T \beta_0)/\sigma_0) X_{1r} \omega(X_1)] = 0, \quad 1 \leq r \leq d, \\
E\{\varphi_r(X_1, Y_1) \gamma_0(Y_1) \delta_1\} &= E[\chi(\omega(X_1)(Y_1 - X_1^T \beta_0)/\sigma_0)] = 0, \quad r = d+1.
\end{aligned}$$

Recalling the definition of  $\Pi$  in (A.7), the Multivariate Central Limit Theorem yields

$$b_n \xrightarrow{d} N(0, \Pi) \quad \text{as } n \rightarrow \infty. \quad (\text{A.9})$$

Let  $A_n = \begin{pmatrix} A_n^{11} & A_n^{12} \\ A_n^{21} & A_n^{22} \end{pmatrix}$ , where

$$\begin{aligned} A_n^{11} &= \frac{\partial^2 S_n(\beta, \sigma)}{\partial \beta \partial \beta^T} \Big|_{\theta=\theta'_n} = \sum_{i=1}^n W_{ni} \psi^{(1)} \left( \frac{\omega(X_i)(Y_i - X_i^T \beta'_n)}{\sigma'_n} \right) \omega^2(X_i) X_i \frac{X_i^T}{\sigma'_n}, \\ A_n^{12} &= \frac{\partial^2 S_n(\beta, \sigma)}{\partial \beta \partial \sigma} \Big|_{\theta=\theta'_n} = \sum_{i=1}^n W_{ni} \psi^{(1)} \left( \frac{\omega(X_i)(Y_i - X_i^T \beta'_n)}{\sigma'_n} \right) \omega^2(X_i) X_i \frac{Y_i - X_i^T \beta'_n}{(\sigma'_n)^2}, \\ A_n^{21} &= \frac{\partial^2 S_n(\beta, \sigma)}{\partial \sigma \partial \beta^T} \Big|_{\theta=\theta'_n} = \sum_{i=1}^n W_{ni} \chi^{(1)} \left( \frac{\omega(X_i)(Y_i - X_i^T \beta'_n)}{\sigma'_n} \right) \omega(X_i) \frac{X_i^T}{\sigma'_n}, \\ A_n^{22} &= \frac{\partial^2 S_n(\beta, \sigma)}{\partial \sigma \partial \sigma} \Big|_{\theta=\theta'_n} = \sum_{i=1}^n W_{ni} \chi^{(1)} \left( \frac{\omega(X_i)(Y_i - X_i^T \beta'_n)}{\sigma'_n} \right) \omega(X_i) \frac{Y_i - X_i^T \beta'_n}{(\sigma'_n)^2}. \end{aligned}$$

Applying the strong law of large numbers in Proposition 1 and the continuity of elements of  $A_n$  about  $\theta$ , we have that  $A_n \rightarrow \Lambda$ , as  $n \rightarrow \infty$ . The assertion of Theorem 2 therefore follows from (A.8) and (A.9).  $\square$

## Appendix B: Detailed Discussions of Remark 2

To start, we assume that the following regularity conditions hold.

(D1)  $T$  and  $C$  are independent condition on  $X$ .

(D2) The bandwidth  $h_n = O(n^{-v})$  with  $0 < v < 1/d$ .

(D3) (i)  $K$  is a kernel with compact support and total variation. (ii) The kernel  $K$  satisfies  $\int x_j K(x) dx = 0$ ,  $j = 1, \dots, d$ .

(D4) (i)  $K$  is a bounded kernel with compact support in  $R^d$ . (ii) The kernel has order  $q$  satisfies  $\int K(x) dx = 1$  and  $\int x_1^{u_1} \dots x_d^{u_d} K(x) dx = 0$  for nonnegative integers  $u_1, \dots, u_d$  with  $\sum_{i=1}^d u_i < q$

(D5) The first  $q$  partial derivatives with respect to  $x$  of the density function  $f(x)$  are uniformly bounded for  $x$ , the conditional density of  $T$  given  $X = x$ ,  $f(t|x)$  and the conditional density of  $C$  given  $X = x$ ,  $g(t|x)$  are uniformly bounded away from infinity and have bounded first  $q$  order partial derivatives with respect  $x$ .

(D6) The bandwidth  $h_n$  satisfies  $h_n = O(n^{-v})$  with  $(2q)^{-1} < v < (3d)^{-1}$ .

Recall that  $\varphi_\theta(x, t) = \rho_0(\omega(t - x^T \beta)/\sigma) + \log(\sigma)$ . The key step to show the consistency is to prove

$$S_n(\theta) = \sum_{i=1}^n \frac{\delta_i}{1 - \hat{G}(Y_i|X_i)} \varphi_\theta(X_i, Y_i) \rightarrow S(\theta) = E[\varphi_\theta(x, t)], \quad \text{a.s.}$$



for any fixed  $\theta \in R^{d+1}$ . Note that

$$\begin{aligned} \sum_{i=1}^n \frac{\delta_i}{1 - \hat{G}(Y_i|X_i)} \varphi_\theta(X_i, Y_i) &= \sum_{i=1}^n \frac{\delta_i}{1 - G(Y_i|X_i)} \varphi_\theta(X_i, Y_i) \\ &+ \sum_{i=1}^n \frac{\delta_i \varphi_\theta(X_i, Y_i) [\hat{G}(Y_i|X_i) - G(Y_i|X_i)]}{[1 - \hat{G}(Y_i|X_i)][1 - G(Y_i|X_i)]}. \end{aligned}$$

By the strong law of large numbers, the first term in the right-hand side of the equation above converges almost surely to  $E[\varphi_\theta(x, t)]$ . Under conditions(D1)-(D3),  $\sup_{y,x} |\hat{G}(y|x) - G(y|x)| = o(1)$  almost surely by Theorem 2.1 in Liang et al. (2012). So the second term can be bounded by  $O(1) \sup_{y,x} |\hat{G}(y|x) - G(y|x)| = o(1)$  almost surely.

To prove the asymptotic normality, using the same notation as in the proof of Theorem 2 in Appendix A, we need to show the asymptotic normality of

$$b_n = -n^{-1/2} \sum_{i=1}^n \frac{\delta_i \varphi(X_i, Y_i)}{1 - \hat{G}(Y_i|X_i)},$$

with vector-valued function  $\varphi(x, y) = [\psi(\omega(y - x^T \beta_0)/\sigma_0) x^T \omega, \chi(\omega(y - x^T \beta_0)/\sigma_0)]^T$ , which is a vector-valued function. Denote  $\varphi_r$  is the  $r$ th component of  $\varphi(\cdot, \cdot)$ . First, using the uniform convergence rate of  $\hat{G}$  given by Theorem 2.1 in Liang et al. (2012), we obtain that under conditions (D1) and (D4)-(D6)

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varphi_r(X_i, Y_i)}{1 - \hat{G}(Y_i|X_i)} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varphi_r(X_i, Y_i) [\hat{G}(Y_i|X_i) - G(Y_i|X_i)]}{[1 - G(Y_i|X_i)]^2} + O_p((\log(n)/(nh_n^d))^{3/4} + h_n^q).$$

Note that the function  $\delta_i \varphi_r(X_i, Y_i) G(Y_i|X_i) [1 - G(Y_i|X_i)]^{-2}$  belongs to a Donsker Class from the assumption and permeance property of the Donsker class (Van der Vaart and Wellner, 1996). It is also the case for  $\delta_i \varphi_r(X_i, Y_i) \hat{G}(Y_i|X_i) [1 - \hat{G}(Y_i|X_i)]^{-2}$  with probability tending to one by Theorem 2.1 in Liang et al. (2012). Applying the asymptotic equicontinuity of the Donsker class (Van der Vaart and Wellner, 1996), we have

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \varphi_r(X_i, Y_i) [\hat{G}(Y_i|X_i) - G(Y_i|X_i)]}{[1 - G(Y_i|X_i)]^2} = \int \frac{\varphi_r(x, y) [\hat{G}(y|x) - G(y|x)]}{1 - G(y|x)} dF(x, y) + o_p(n^{-1/2}),$$

uniformly for  $x, y$  and  $\varphi_r$ . By Theorem 2.3 in Liang et al. (2012), we get

$$\begin{aligned} &\int \frac{\varphi_r(x, y) [\hat{G}(y|x) - G(y|x)]}{1 - G(y|x)} dF(x, y) \\ &= \sum_{i=1}^n \int \frac{K(\frac{X_i - x}{h_n})}{\sum_{j=1}^n K(\frac{X_j - x}{h_n})} \varphi_r(x, y) \xi(Y_i, \delta_i, y, x) dF(x, y) + O_p((\log(n)/(nh_n^d))^{3/4} + h_n^q), \end{aligned}$$

where  $\xi(Y, \delta, y, x) = \frac{I(Y \leq y, \delta = 0)}{C(Y|x)} - \int_0^y \frac{I(t \leq Y)}{C^2(t|x)} dH_1(t|x)$ ,  $C(y|x) = P(Y \geq y|X = x)$  and  $H_1(y|x) = P(Y \leq y, \delta = 0|X = x)$ .

Let  $\hat{f}(x) = (nh_n^d)^{-1} \sum_{j=1}^n K([X_j - x]/h_n)$ , which can be decomposed into three parts:

$$\begin{aligned} & \frac{1}{nh_n^d} \sum_{i=1}^n \int K\left(\frac{X_i - x}{h_n}\right) \frac{\varphi_r(x, y) \xi(Y_i, \delta_i, y, x)}{f(x)} dF(x, y) \\ & + \frac{1}{nh_n^d} \sum_{i=1}^n \int K\left(\frac{X_i - x}{h_n}\right) \frac{\varphi_r(x, y) \xi(Y_i, \delta_i, y, x) [\hat{f}(x) - f(x)]}{f(x)^2} dF(x, y) \\ & + \frac{1}{nh_n^d} \sum_{i=1}^n \int K\left(\frac{X_i - x}{h_n}\right) \frac{\varphi_r(x, y) \xi(Y_i, \delta_i, y, x) [\hat{f}(x) - f(x)]^2}{\hat{f}(x) f(x)^2} dF(x, y), \end{aligned}$$

where  $F(x, y)$  is the joint distribution of  $X$  and  $Y$ .

By a change of variables and the second order Taylor expansion, the first term in the expression above can be written as  $n^{-1} \sum_{i=1}^n \int \varphi_r(X_i, y) \xi(Y_i, \delta_i, y, X_i) dF_{X_i}(x, y) + O_p(h_n^2)$  with  $F_X(x, y) = P(X \leq x, Y \leq y|X = x)$ , while the second term can be written as

$$\int \varphi_r(x, y) \left\{ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \frac{\xi(Y_i, \delta_i, y, x)}{f(x)^2} \right\} \left\{ \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) - f(x) \right\} dF(x, y),$$

and consequently equals to  $O_p(\log(n)(nh_n^d)^{-1} + h_n^{2q})$  using the uniform convergence rate of  $\hat{f}(x)$  (Theorem 6 in Hanson, 2008). Similarly, it is easy to see that the third term is  $O_p(\log(n)(nh_n^d)^{-1} + h_n^{2q})$  uniformly over  $\psi_r$ . This establishes the desired asymptotic normality by the Central Limit Theorem. Therefore, results of both Theorems 1 and 2 can be applied under the weak condition addressed in Remark 2.

## References

- [1] Bianco, A. M., Ben, M. G., and Yohai, V. J. (2009). Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics* **33**, 511–528.
- [2] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.
- [3] Cui, H.J. (2004). On asymptotic of t-type regression estimation in multiple linear model. *Science in China Ser. A.* **34**, 361–372.
- [4] Gonzalez-Manteiga, W., Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics* **4**, 65–78.

- [5] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.
- [6] Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.
- [7] He, X. and Fung, W.K. (1999). Method of medians for lifetime data with Weibull models. *Statistics in Medicine* **18**, 1993–2009.
- [8] He, X., Cui, H. and Simpson, D. (2004). Longitudinal data analysis using t-type regression. *Journal of Statistical Planning and Inference* **122**, 253–269.
- [9] He, X., Simpson, D. and Wang G.Y. (2000). Breakdown points of t-type regression estimators. *Biometrika* **87**, 675–687.
- [10] Honore, B., S. Khan, and J. Powell (2002). Quantile regression under random censoring. *Journal of Econometrics* **64**, 241–278.
- [11] Huang, J. Ma, S., and Xie, H. (2007). Least absolute Deviations Estimation For The Accelerated Failure Time Model. *Statistica Sinica* **17**, 1533–1548.
- [12] Huber, P. J. (1981). *Robust Statistics*. John Wiley, New York.
- [13] Jin, Z. (2007). M-estimation in regression models for censored data. *Journal of Statistical Planning and inference* **137**, 3894–3903.
- [14] Jin, Z., Lin, D.Y., Wei, L.J. and Jing, Z. (2003). Rank-based inference for the accelerated Failure Time Model. *Biometrika* **90**, 341–353.
- [15] Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd ed. Wiley, New York.
- [16] Komarek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association* **103**, 523–533.
- [17] Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded influence regression estimation. *Journal of the American Statistical Association* **77**, 595–604.

- [18] Lai, T.L. and Ying, Z.L. (1994). A missing information principle and M-estimator in regression analysis with censored and truncated data. *Annals of Statistics* **22**, 1222–1225.
- [19] Liang, H.Y., de Uña-Álvarez and Iglesias-Pérez, M. (2012). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Test* **21**, 790–810.
- [20] Locatelli, I., Marazzi, A. and Yohai, V.J. (2010). Robust accelerated failure time regression. *Computational Statistics and Data Analysis* **55**, 874–887.
- [21] Lopez, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Communications in Statistics-Theory and Methods* **40**, 2639–2660.
- [22] Ma, S. and Kosorok, M.R. (2005). Robust semiparametric M-estimation and weighted bootstrap. *Journal of Multivariate Analysis* **96**, 190–217.
- [23] Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, New York..
- [24] Peng, L. and Huang, Y. (2008). Survival Analysis With Quantile Regression Models. *Journal of American Statistical Association* **103**, 637–649.
- [25] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7**, 186–199.
- [26] Portnoy, S. (2003). Censored Regression Quantiles. *Journal of American Statistical Association* **98**, 1001–1012.
- [27] Reid, N. (1981). Influence Functions for Censored Data. *Annals of Statistics* **9**, 78–92.
- [28] Richardson, G.D. and Bhattacharyya, B.B. (1986). Consistent estimators in nonlinear regression for a noncompact parameter space. *Annals of Statistics* **14**, 1591–1596.
- [29] Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics* **18**, 303–328.
- [30] Rousseeuw, P.J. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223.

- [31] Salibian-Barrera, M. and Yohai, V.J. (2008). High breakdown point robust regression with censored data. *Annals of Statistics* **36**, 118–146.
- [32] Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*. John Wiley, New York.
- [33] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89–103.
- [34] Stute, W. (1995). The central limit theorem under random censorship. *Annals of Statistics* **23**, 422–439.
- [35] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics* **23**, 461–471.
- [36] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.
- [37] Stute, W. and Wang, J.L. (1993). The strong law under random censorship. *Annals of Statistics* **21**, 1591–1607.
- [38] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.
- [39] Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **104**, 1117–1128.
- [40] Ying, Z.L. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* **21**, 76–99.
- [41] Ying, Z.L., Jung, S.H. and Wei, L.J. (1995). Survival analysis with median regression models. *Journal of Statistical Planning and Inference* **90**, 178–184.
- [42] Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837–841.
- [43] Zhou, W. (2010). M-estimation of the accelerated failure time model under a convex discrepancy function. *Journal of Statistical Planning and Inference* **140**, 2669–2681.

Table 1. Simulation results: means (root mean square errors) of the proposed KMW-GM estimates in comparison with the KMW-LAD estimates,  $(\beta_1, \beta_2) = (0, 1)$  and  $\epsilon \sim N(0, 1)$ .

sample size	censoring rate	KMW-LAD		KMW-GM	
		$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
100	0%	-0.018	0.998	-0.015	0.999
		(0.257)	(0.075)	(0.214)	(0.062)
	10%	0.010	0.994	0.004	0.995
		(0.246)	(0.074)	(0.209)	(0.062)
	40%	0.026	1.014	0.039	1.009
		(0.303)	(0.096)	(0.266)	(0.080)
500	0%	-0.005	1.000	-0.006	1.001
		(0.107)	(0.032)	(0.092)	(0.027)
	10%	-0.002	1.000	-0.001	1.000
		(0.118)	(0.034)	(0.097)	(0.028)
	40%	0.012	0.998	0.021	0.994
		(0.143)	(0.048)	(0.122)	(0.040)

Table 2. Simulation results: means (root mean square errors) of the proposed KMW-GM estimates in comparison with the KMW-LAD estimates,  $(\beta_1, \beta_2) = (0, 1)$  and  $\epsilon \sim t(3)/\sqrt{3}$ .

sample size	censoring rate	KMW-LAD		KMW-GM	
		$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
100	0%	-0.006	0.998	-0.008	0.999
		(0.157)	(0.046)	(0.147)	(0.043)
	10%	-0.003	0.999	-0.002	0.999
		(0.168)	(0.050)	(0.157)	(0.046)
	40%	0.014	1.002	0.023	1.000
		(0.204)	(0.068)	(0.197)	(0.062)
500	0%	0.001	0.999	-0.001	0.999
		(0.071)	(0.021)	(0.066)	(0.019)
	10%	-0.001	1.000	-0.001	1.000
		(0.072)	(0.021)	(0.097)	(0.020)
	40%	0.004	0.999	0.010	0.997
		(0.086)	(0.030)	(0.083)	(0.027)

Table 3. Simulation results: root mean square error under the standard Log-Weibull error

Parameter	Estimate	Sample size			
		100	200	300	500
$\beta_1$	KMW-LAD	0.235	0.228	0.216	0.203
	KMW-GM	0.209	0.185	0.172	0.159
$\beta_2$	KMW-LAD	0.201	0.139	0.118	0.097
	KMW-GM	0.212	0.142	0.126	0.105

Table 4(I). Simulation results: root mean square errors of slope  $\beta_2$  with 10% outliers at  $(x_0, mx_0)$ ,

$x_0 = 10$				
Estimator	$m$			
	2	3	4	5
KMW-LAD	0.502	1.271	1.793	2.296
KMW-GM	0.155	0.184	0.270	0.331

Table 4(II). Simulation results: means (root mean square errors) of the proposed KMW-GM estimates in comparison with the KMW-LAD estimates,  $(\beta_1, \beta_2) = (0, 1)$ ,

$\epsilon_i \sim 0.9N(0, 1) + 0.1N(0, 16)$  and  $x_i \sim 0.9N(0, 1) + 0.1N(-3, 1/9)$ .

Censoring rate	KMW-LAD		KMW-GM	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
0%	0.0173	0.9015	0.0034	0.9257
	(0.1352)	(0.1667)	(0.1190)	(0.1523)
10%	-0.0174	0.8738	-0.0307	0.8916
	(0.1365)	(0.2046)	(0.1270)	(0.1836)
40%	0.0480	0.9267	0.0307	0.9215
	(0.2153)	(0.2617)	(0.2079)	(0.2198)



Table 5. Simulation results: standard errors computed by the weighted bootstrap (sample standard deviations) of the KMW-GM estimates with  $(\beta_1, \beta_2) = (0, 1)$  under different error distributions.

sample size		Error Distribution		
		N(0,1)	$t(3)/\sqrt{3}$	$t(5)/\sqrt{5/3}$
100	$\beta_1$	0.276(0.263)	0.199(0.196)	0.234(0.239)
	$\beta_2$	0.081(0.079)	0.084(0.062)	0.071(0.073)
500	$\beta_1$	0.116(0.120)	0.088(0.083)	0.106(0.102)
	$\beta_2$	0.042(0.039)	0.025(0.027)	0.037(0.033)

Table 6. Simulation results: means (root mean square errors) of the KMW-GM estimates with  $(\beta_1, \beta_2) = (0.5, 0.3)$  and  $N(0, 1)$  errors.

sample size	censoring rate	$\beta_1$	$\beta_2$
100	0%	0.499(0.020)	0.299(0.022)
	10%	0.496(0.024)	0.298(0.025)
	40%	0.491(0.038)	0.291(0.040)
300	0%	0.500(0.012)	0.299(0.012)
	10%	0.500(0.014)	0.299(0.015)
	40%	0.494(0.021)	0.295(0.022)

Table 7. Simulation results: means (root mean square errors) of the KMW-GM estimates with  $(\beta_1, \beta_2) = (0.5, 0.3)$  and  $t(3)/\sqrt{3}$  errors.

sample size	censoring rate	$\beta_1$	$\beta_2$
100	0%	0.499(0.014)	0.300(0.016)
	10%	0.497(0.017)	0.299(0.019)
	40%	0.495(0.026)	0.296(0.026)
300	0%	0.500(0.008)	0.300(0.009)
	10%	0.500(0.010)	0.300(0.010)
	40%	0.497(0.015)	0.298(0.016)

Table 8. Simulation results: means (root mean square errors) of the KMW-GM estimates with  $(\beta_1, \beta_2) = (0.5, 0.3)$  and  $t(5)/\sqrt{5/3}$  errors.

sample size	censoring rate	$\beta_1$	$\beta_2$
100	0%	0.499(0.017)	0.300(0.020)
	10%	0.496(0.021)	0.298(0.021)
	40%	0.493(0.033)	0.295(0.034)
300	0%	0.499(0.010)	0.300(0.011)
	10%	0.499(0.011)	0.300(0.012)
	40%	0.496(0.018)	0.295(0.020)

Table 9. Simulation results: means (root mean square errors) of estimates obtained by the proposed method with the local K-M estimate weight (KMW-GM) in comparison with censored quantile regression methods developed by Portnoy(2003) (CRQP), Peng and Huang(2008) (CRQPH) and Wang and Wang (2009) (LCRQ).  $(\beta_1, \beta_2) = (3, 5)$ .

Sample size	Method	$N(0, 1)$		$t(3)/\sqrt{3}$		$Laplace(0, \sqrt{0.5})$	
		$b_0$	$b_1$	$b_0$	$b_1$	$b_0$	$b_1$
200	LCRQ	2.992	4.980	3.001	4.999	2.992	5.004
		(0.217)	(0.388)	(0.136)	(0.253)	(0.132)	(0.253)
	CRQP	2.958	4.995	2.976	5.005	2.969	5.012
		(0.221)	(0.389)	(0.135)	(0.248)	(0.136)	(0.253)
	CRQPH	3.018	4.997	3.016	5.001	3.004	5.014
		(0.217)	(0.389)	(0.135)	(0.250)	(0.131)	(0.250)
KMW-GM	3.006	4.979	3.007	4.990	2.993	5.006	
	(0.184)	(0.330)	(0.130)	(0.244)	(0.149)	(0.278)	
500	LCRQ	2.979	5.013	2.992	5.001	2.996	4.993
		(0.129)	(0.246)	(0.086)	(0.160)	(0.082)	(0.150)
	CRQP	2.957	5.024	2.978	5.004	2.981	4.997
		(0.133)	(0.246)	(0.087)	(0.159)	(0.083)	(0.147)
	CRQPH	2.998	5.031	3.003	5.007	3.005	5.002
		(0.128)	(0.251)	(0.085)	(0.159)	(0.081)	(0.151)
KMW-GM	2.991	5.011	3.000	4.999	3.004	4.985	
	(0.116)	(0.213)	(0.081)	(0.153)	(0.095)	(0.178)	

Table 10. Results of analysis of the Heart data under the AFT model.

parameter	KMW-LAD		KMW-GM	
	estimate (s.e.)	p-value	estimate (s.e.)	p-value
Complete data				
$\alpha$	7.827 (1.841)	0.001	6.275(1.174)	0.000
$\beta_1$	-0.057 (0.039)	0.148	-0.018(0.025)	0.482
$\beta_2$	0.462 (0.669)	0.491	0.266(0.453)	0.558
$\sigma$	1.896 (0.099)		1.334(0.080)	
Outliers removed				
$\alpha$	7.826 (1.934)	0.001	6.371(1.210)	0.000
$\beta_1$	-0.057 (0.042)	0.159	-0.018(0.026)	0.492
$\beta_2$	0.460 (0.717)	0.484	0.245(0.442)	0.580
$\sigma$	1.649 (0.107)		1.257(0.066)	

Note: s.e.= standard error.

Table 11. The relative variation of the estimates with or without outliers for the Heart data.

parameter	KMW-LAD	KMW-GM
$\alpha$	0.0001582485	0.015262402
$\beta_1$	0.0028956305	0.003423491
$\beta_2$	0.0054283995	0.080651249
$\sigma$	0.1493182265	0.057279113

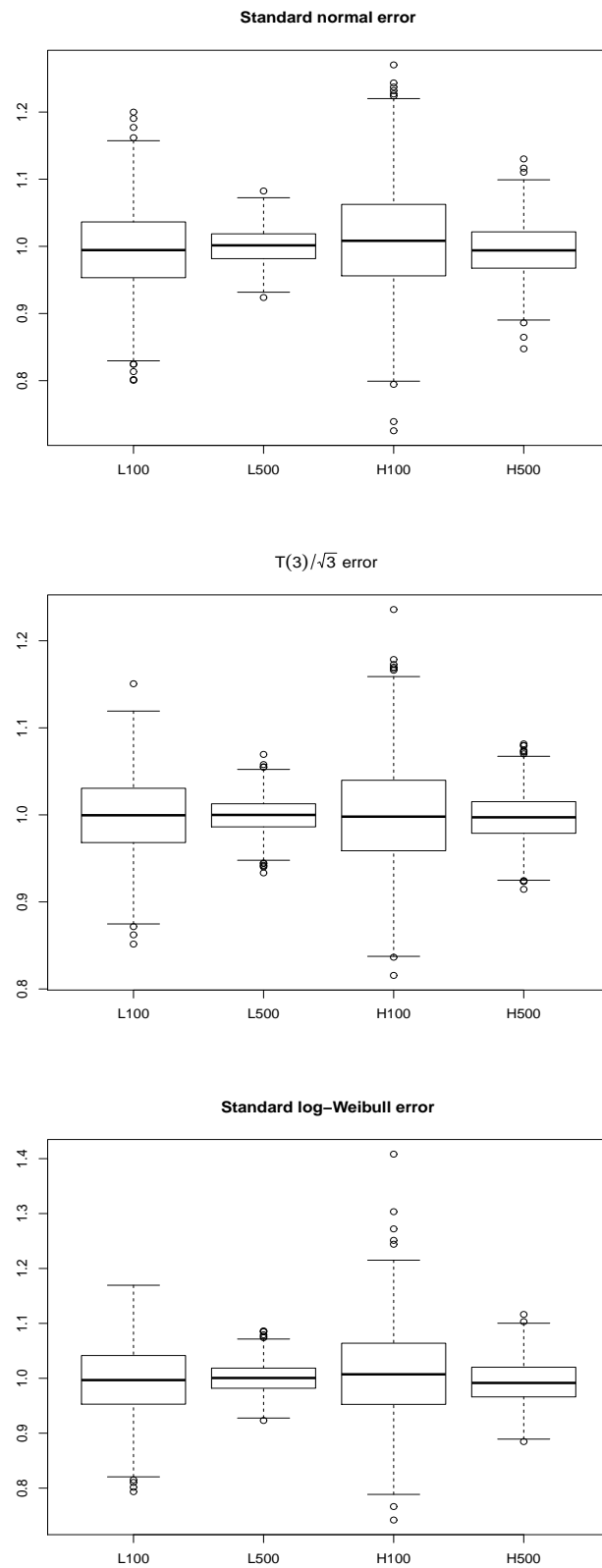


Fig 1. Boxplots of estimates of  $\beta_2$  with differently distributed errors for censoring rates of 10% (L) and 40% (H), and  $n = (100, 500)$ .

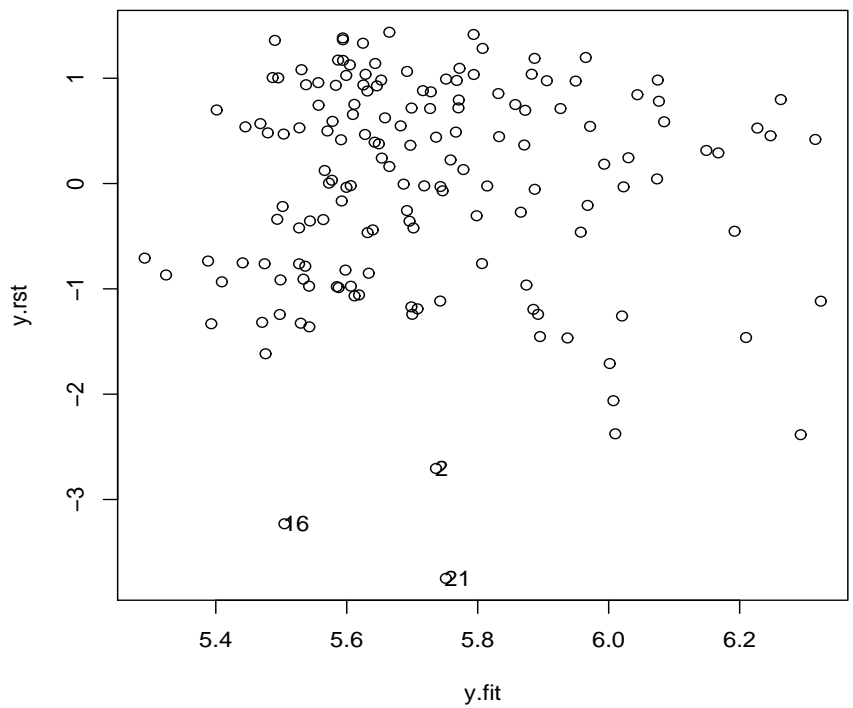
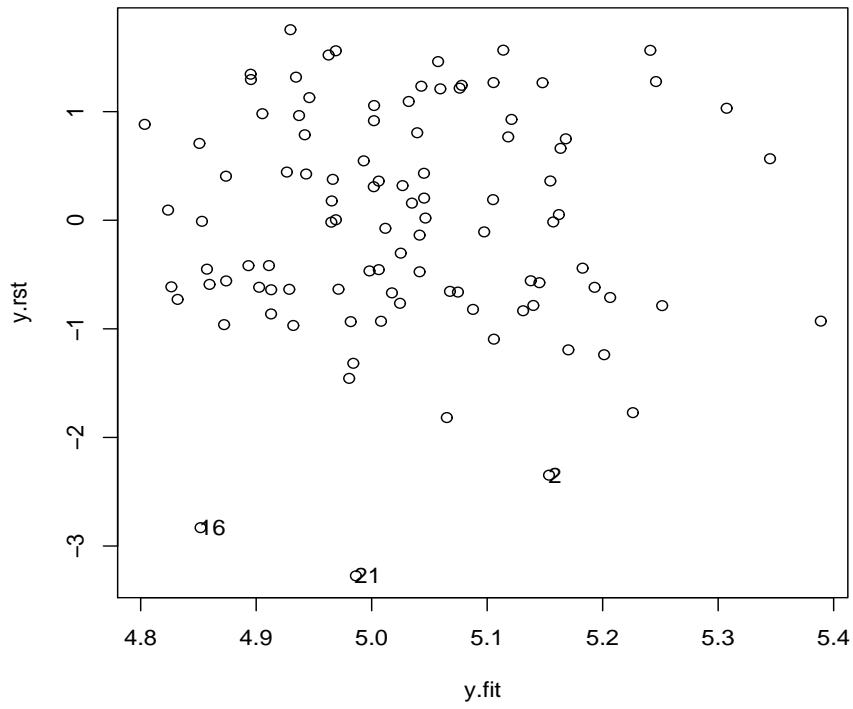


Figure 2. Standard residuals corresponding to the uncensored data set (top) and the complete data set (bottom).