Statistica Sinica Preprint No: SS-2023-0229	
Title	Heteroscedastic Survival Data Analysis with Accelerated Failure Time Model
Manuscript ID	SS-2023-0229
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0229
Complete List of Authors	Lili Yu and Zhezhen Jin
Corresponding Authors	Lili Yu
E-mails	lyu@georgiasouthern.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Heteroscedastic survival data analysis with accelerated failure time model

Lili Yu[1] and Zhezhen Jin[2]

Georgia Southern University[1] *and Columbia University*[2]

Abstract: Recently, the accelerated failure time model has been extended to accommodate heteroscedastic survival data. However, the existing methods often require stringent assumptions or complex algorithms. In this paper, a weighted least squares method is developed based on Laplace approximation for quasi-likelihood subject to conditionally independent censoring. The Laplace approximation is used to approximate the quasi-survival function of the censored observations, which results in simpler and more computationally efficient estimation than the existing methods. The consistency and asymptotic distribution of the resulting estimator are also established. Extensive simulations are conducted to evaluate the performance of the proposed method. Finally, we apply the new proposed method to Stanford heart transplant data and colon cancer data to demonstrate its use in real applications.

Key words and phrases: Heteroscedasticity; Laplace approximation; Local polynomial regression.

1. Introduction

Right censored survival data is the most popular data type in survival analysis. It is characterized by the possibility of censoring time, which refers to observations that terminate before the events of interest can be observed. This may be due to the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

reasons such as being alive at the end of study, dying from other reasons or losing contact before the end for various reasons. The Cox model (Cox (1972)) is the most popular model for analyzing such data. However, the accelerated failure time (AFT) model (Wei, Ying and Lin (1990); Kalbfleisch and Prentice (2002)) provides an attractive alternative because it directly interprets the effects of covariates on the mean survival time. It was traditionally proposed to handle homoscedastic survival data. Many inference methods have been proposed for the homoscedastic AFT model. Under the strong unconditional independence assumption that survival time and censoring time are unconditionally independent, Koul, Susarla and Van Ryzin (1981), Leurgans (1987), and Fan and Gijbels (1994), proposed to replace the censored observations by synthetic data constructed by inverse weighted probability of censoring distribution. Under a weaker conditional independence assumption (currently considered as the usual assumption) that the survival time and censoring time are independent conditional on covariates, three main methods have been studied in the literature: the rank-based method (Tsiatis (1990); Lai and Ying (1991b); Robins and Tsiatis (1992); Ying (1993); Lin and Ying (1995); Jin et al. (2003); Zhou (2005)), the least squares method (Buckley and James (1979); Ritov (1990); Lai and Ying (1991a); Jin, Lin and Ying (2006)) and the profile likelihood method (Zeng and Lin (2007)).

On the other hand, it is challenging to use the AFT model for the analysis of heteroscedastic survival data, although the heteroscedastic survival data are often seen in real applications, such as the colon cancer data in Section 6 of this paper. It aims

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

to compare the effects of three different treatments on the survival time for patients with colon cancer. The variance plot shown in Section 6 indicates that the variance of the colon cancer data is not homoscedastic, but is a function of the data mean. Stare, Heinzl and Harrell (2000) found that the least squares estimator is biased for heteroscedastic survival data. Several researchers have attempted to address the issue: see Zhang and Davidian (2008), Chen and Khan (2000), Zhou, Bathke and Kim (2012), Heuchenne and Van Keilegom (2007). Some of the estimation methods for the homoscedastic AFT model have been extended to heteroscedastic AFT model. Under the strong unconditional independence assumption that the survival time and censoring time are unconditionally independent, Liu and Lu (2009) extended the synthetic data approach via inverse probability weighting method (Koul, Susarla and Van Ryzin (1981); Leurgans (1987); Fan and Gijbels (1994)) by handling the heteroscedasticity with a kernel smoothing approach. Under the weaker conditional independence assumption that the survival time and censoring time are independent conditional on covariates, the least squares method has been extended to the analysis of heteroscedastic survival data in two ways. Yu (2011), Yu et al. (2012), Yu and Peace (2012), Yu, Liu and Chen (2013) proposed weighted least squares approaches by handling the heteroscedasticity with various nonparametric estimation methods. Pang, Lu and Wang (2015) modified Buckley and James' approach using locally estimated Kaplan-Meier survival function to construct synthetic data and accommodate the heteroscedasticity simultaneously. However, the two approaches (Yu, Liu and Chen (2013); Pang, Lu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

and Wang (2015)) have some drawbacks. First, the resulting estimating equations are neither continuous nor monotone in the parameter of interest, which results in computational difficulty. Both approaches involve Kaplan-Meier estimation in each iteration, which is time-consuming and sometimes yields unstable estimates. As a result, they are often not feasible for the analysis of survival data with high censoring or large sample size. In addition, it hampers further extensions to more complex settings, such as heteroscedastic survival data with cure or frailty. Second, the variance function and the synthetic data are "bundled together" (Ding and Nan (2011)), i.e., the variance function and the synthetic data depend on each other. Therefore, misspecification of the variance function will affect the synthetic data and hence the parameter estimation.

In this paper, we develop a novel weighted least squares method based on Laplace approximation for quasi-likelihood subject to the assumption of conditional independence between the survival time and censoring time. Our approach uses Laplace approximation to approximate the quasi-survival function of the censored observations, instead of constructing the synthetic data. It overcomes the challenges in the existing methodologies. First, it results in a continuous and monotone estimating equation, which facilitates the estimation procedure. The approach does not require the calculation of the Kaplan-Meier estimator iteratively, which greatly simplifies computation and facilitates its extension to more complex settings. Second, there is no close dependence between Laplace approximated observations and the variance function.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Therefore, the parameter estimation is robust to variance function estimation.

This paper is organized as follows. Section 2 describes the weighted least squares method based on Laplace approximation for quasi-likelihood. Section 3 presents the asymptotic properties. Section 4 provides the bias of the proposed estimator and presents a detailed algorithm on the implementation of the proposed method. Section 5 reports simulation results and section 6 presents real data analyses. Section 7 provides concluding remarks and discussions. The online supplementary material sketches the proofs of the asymptotic properties and presents additional simulation results.

2. The Laplace approximated weighted least squares method

2.1 Data and Model Setup

Let T_i be the survival time and Y_i, C_i be the logarithm of survival time and censoring time, respectively. Due to right censoring, we observe $\{y_i, \delta_i, \mathbf{X}_i\}, i = 1, \cdots, n$, where $y_i = \min(Y_i, C_i)$ is the observed logarithm of survival time, $\delta_i = I(Y_i \leq C_i)$ is the censoring indicator and $\mathbf{X}_i$ is a p-vector of covariates with the first element being one for the ith observation. It is assumed that Y_i and C_i are independent conditional on $\mathbf{X}_i$.

To accommodate the heteroscedasticity in datasets such as the colon cancer data in Section 6, we adopt the following heteroscedastic AFT model (Yu, Liu and Chen (2013); Pang, Lu and Wang (2015)),

$$Y_i = \mu_i^* + \sigma^*(\mu_i^*)e_i^*, \tag{2.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

where $\mu_i^* = \boldsymbol{\beta}^{*T}\mathbf{X}_i$ with $\boldsymbol{\beta}^*$ being a p-vector of unknown parameters; $\sigma^*(\mu_i^*)$ is the square root of the variance of Y_i, which accounts for the heteroscedasticity of the data; e_i^* is a random error with mean zero and variance one.

2.2 The Laplace approximated weighted least squares method

When there is no censoring, the classical weighted least squares method estimates the parameter $\boldsymbol{\beta}^*$ by solving

$$\sum_{i=1}^{n} \mathbf{X}_i^T \sigma^{*(-2)}(\mu_i^*)(Y_i - \mu_i^*) = 0.$$

It is equivalent to maximizing the quasi-likelihood (Wedderburn (1974); Heyde (1997)),

$$\sum_{i=1}^{n} \int_{Y_i}^{\mu_i^*} (Y_i - a)/\sigma^{*2}(a)da.$$

Quasi-likelihood is an extension of the generalized linear model. It only requires assumptions on the first two moments which is simpler to specify than assuming the full likelihood. It is identical to the true log likelihood when the true distribution belongs to the exponential family. Because the quasi-likelihood has similar algebraic and frequency properties of a log likelihood function, that is, the quasi-score function (derivative of the quasi-likelihood) behaves like the derivative of a log likelihood (Heyde (1997)), we can construct a corresponding likelihood function and the parameter can be estimated by maximizing the likelihood function.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Now define the quasi-density function for the variable Y as

$$f(y|\mathbf{X};\boldsymbol{\beta}^*) = h(y)\exp\{\int_y^{\mu^*}(y-a)/\sigma^{*2}(a)da\},$$

and the corresponding quasi-survival function

$$S(y|\mathbf{X};\boldsymbol{\beta}^*) = \int_y^{\infty} h(u)\exp\{\int_u^{\mu^*}(u-a)/\sigma^{*2}(a)da\}du,$$

where $h(y) \geq 0$ is some function of y, which makes the quasi-density to be a probability density function, that is,

$$\int f(y|\mathbf{X};\boldsymbol{\beta}^*)dy = 1.$$

For example, let $q(y)$ be some function of y. Then

$$h(y) = \frac{q(y)}{\int q(y)\exp\{\int_y^{\mu^*}(y-a)/\sigma^{*2}(a)da\}dy},$$

where $\int q(y)\exp\{\int_y^{\mu^*}(y-a)/\sigma^{*2}(a)da\}dy$ is the normalizing term. This normalizing term may depend on the unknown parameter $\boldsymbol{\beta}^*$, so we set it as its true value and then $h(y)$ is a function only for y. Because the quasi-density is a probability density function, the quasi-survival function is $p(Y > y|\mathbf{X};\boldsymbol{\beta}^*)$ and it then can be used to construct the likelihood for censored observations. Note that the derivative of the log quasi-density function is equivalent to the quasi-score function. Therefore, the maximum quasi-likelihood estimator is the same as the estimator that maximizes the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

likelihood function constructed based on the quasi-density. We call the log likelihood constructed based on the quasi-density as quasi-density-likelihood.

In the presence of the censoring, we construct the quasi-density-likelihood using the quasi-survival function for censored observations, instead of constructing synthetic observations as in Yu, Liu and Chen (2013) and Pang, Lu and Wang (2015). Specifically, the quasi-density-likelihood for the censored data is

$$\mathrm{l}(\boldsymbol{\beta}^*) = \sum_{i=1}^{n} \left[\delta_i \left\{ \log(h(y_i)) + \int_{y_i}^{\mu_i^*} \frac{(y_i - a)}{\sigma^{*2}(a)} da \right\} + (1 - \delta_i) \log \left\{ \int_{y_i}^{\infty} h(y_i) \exp\left(\int_{y_i}^{\mu_i^*} \frac{(y_i - a)}{\sigma^{*2}(a)} da \right) dy_i \right\} \right].$$

The quasi-survival function in the quasi-density-likelihood may not be possible to calculate directly because it involves an unknown function $h(.)$, which may require the value of the unknown parameter. We propose to use a Laplace approximation for the quasi-survival function. Fortunately, the final estimating equation is free of $h(.)$ function due to the Laplace approximation.

The Laplace approximation was proposed by Laplace (1774) to approximate the integrals of the form

$$g(w) = \int_{t \in D} \exp(-wr(t)) h(t) dt,$$

where $r(t) > 0$ and $w \to \infty$. By Taylor expansion of $r(t)$ and approximating it to quadratic order, the $g(w)$ can be approximated by

$$g(w) = \frac{\exp(-w\tilde{r})\tilde{h}\sqrt{2\pi}}{\sqrt{w}\sqrt{\tilde{r}''}}(1 + O(w^{-1})),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

where $\tilde{r} = r(\tilde{t})$, $\tilde{r}'' = r''(\tilde{t})$, $\tilde{h} = h(\tilde{t})$ and $\tilde{t} \in D$ is the value of t that minimizes $r(t)$.

For quasi-survival function $S(y|\mathbf{X};\boldsymbol{\beta}^*)$, let $r(t,\boldsymbol{\beta}^*) = \int_{\mu^*}^{t}(t-a)/\sigma^{*2}(a)da$, so $w = 1$. Researchers (Breslow and Clayton (1993); Butler and Wood (2002); Harding and Hausman (2011)) demonstrated that the Laplace approximation performs very well when $w = 1$ and even in subasymptotic cases where w remains small. Algebra shows that the Laplace approximated quasi-survival function is

$$S(y|\mathbf{X};\boldsymbol{\beta}^*) \approx \exp(-r(\tilde{y},\boldsymbol{\beta}^*))(2\pi\sigma^{*2}(\tilde{y}))^{-1/2}h(\tilde{y}),$$

where $r(y,\boldsymbol{\beta}^*) = \int_{\mu^*}^{y}(y-a)/\sigma^{*2}(a)da$, $\tilde{y} = \mu^*$ if $y \le \mu^*$; $\tilde{y} = y$ if $y > \mu^*$, i.e., $\tilde{y} = \max(\mu^*, y)$. The corresponding Laplace approximated quasi-density-likelihood is

$$\tilde{l}(\boldsymbol{\beta}^*) = \sum_{i=1}^{n}\{\delta_i(\log(h(y_i)) - r(\tilde{y}_i,\boldsymbol{\beta}^*)) - (1-\delta_i)\left(r(\tilde{y}_i,\boldsymbol{\beta}^*) + 1/2\log(2\pi\sigma^{*2}(\tilde{y}_i)) - \log(h(\tilde{y}_i))\right)\},$$

where $\tilde{y}_i = (\delta_i + (1-\delta_i)\lambda_i^*)y_i + (1-\delta_i)(1-\lambda_i^*)\mu_i^*$ and $\lambda_i^* = I(y_i > \mu_i^*)$. Note that when $\lambda_i^* = 0$, the observation has value $\mu_i^* = max\{\mu_i^*, y_i\}$, which is unknown. Therefore, it has no contributions to the maximum Laplace approximated quasi-density-likelihood estimator and then maximizing $\tilde{l}(\boldsymbol{\beta})$ is equivalent to maximizing

$$\sum_{i=1}^{n}\{\delta_i(\log(h(y_i)) - r(\tilde{y}_i,\boldsymbol{\beta}^*)) - (1-\delta_i)\lambda_i^*\left(r(\tilde{y}_i,\boldsymbol{\beta}^*) + 1/2\log(2\pi\sigma^{*2}(\tilde{y}_i)) - \log(h(\tilde{y}_i))\right)\}.$$

The estimating equation based on the Laplace approximated quasi-density-likelihood

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

is,

$$\tilde{\mathbf{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\tilde{y}_i - \mu_i)\mathbf{X}_i/\sigma_n^2(\mu_i), \tag{2.2}$$

where $\sigma_n^2(.)$ is a proper estimator of $\sigma^2(.)$, the variance of the Laplace approximated observations $\tilde{y}_i$. The $\mu_i = \boldsymbol{\beta}^T\mathbf{X}_i$ are the means of the Laplace approximated observations. Note that because we use the Laplace approximated observations, the mean and variance in the equation (2.2) is actually for the Laplace approximated observations. When $\lambda_i \equiv I(y_i > \mu_i) = 0$, it does not affect the value of $\tilde{\mathbf{U}}(\boldsymbol{\beta})$, i.e., the $\tilde{\mathbf{U}}_i(\boldsymbol{\beta}) = (\tilde{y}_i - \mu_i)\mathbf{X}_i/\sigma_n^2(\mu_i) = 0$. Because the $\boldsymbol{\beta}$ and $\sigma^2(.)$ are parameters for the Laplace approximated observations, they may not be the same as $\boldsymbol{\beta}^*$ and $\sigma^{*2}(.)$, respectively. Now we describe how to estimate $\boldsymbol{\beta}$ and then how to estimate $\boldsymbol{\beta}^*$ from $\boldsymbol{\beta}$. Note that we use $\boldsymbol{\beta}^*$ to represent the parameters for original data Y and use $\boldsymbol{\beta}$ as the parameters for Laplace approximated data $\tilde{y}_i$. Later in this paper, we use $\boldsymbol{\beta}_0^*$ and $\boldsymbol{\beta}_0$ as the true values for $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}$, respectively.

The Laplace approximated weighted least square estimate $\tilde{\boldsymbol{\beta}}$ can be obtained by solving $\tilde{\mathbf{U}}(\boldsymbol{\beta}) = 0$. A proper estimator of $\sigma^2(.)$, $\sigma_n^2(.)$ in the estimating equation (2.2) can be obtained by the use of the local polynomial regression approach (Fan and Gijbels (1996)). Specifically, we use the following model,

$$\epsilon_i^2 = \sigma^2(\mu_i) + \tau_i,$$

where $\epsilon_i^2 = (\tilde{y}_i - \mu_i)^2$ is the observed value of $\sigma^2(\mu_i)$ and τ_i is the random error with

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

mean 0. Note that in this model, we only assume that the variance function $\sigma^2(\mu)$ depends on μ which is explained in model (2.1), but the form of the variance function is completely unspecified. Adopting the approach in Fan and Gijbels (1996), the variance function estimator can be obtained by

$$\sigma_n^2(u) = \sum_{i=1}^{n} a_i(u)\epsilon_i^2,$$

where

$$a_i(u) = \frac{(nb)^{-1}K(\frac{\mu_i-u}{b})\{A_{n,2}(u) - (\mu_i - u)A_{n,1}(u)\}}{A_{n,0}(u)A_{n,2}(u) - A_{n,1}^2(u)},$$

and $A_{n,j}(u) = (nb)^{-1}\sum_{i=1}^{n} K(\frac{\mu_i-u}{b})(\mu_i - u)^j, j = 0, 1, 2$ with bandwidth b and a kernel function $K(.)$. To avoid boundary effects as in Chiou and Muller (1999), we only consider $\mu \in I$ where I is a compact interval and contained in the interior of the support of f_μ, the density of $\mu = \boldsymbol{\beta}^T\mathbf{X}$, which varies with $\mathbf{X}$.

We will use iterative algorithm, which is introduced in Section 4.3 below, to solve the equation (2.2). In the iterative procedure, the variance is estimated by

$$\tilde{\sigma}_n^2(u) = \sum_{i=1}^{n} \tilde{a}_i(u)\tilde{\epsilon}_i^2,$$

where $\tilde{a}_i(u)$ and $\tilde{\epsilon}_i$ are $a_i(u)$ and ϵ_i evaluated at the estimator $\tilde{\boldsymbol{\beta}}$.

3 Properties of the estimator

When λ and the variance function $\sigma^2(.)$ are known, the maximum likelihood the-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

ory shows that the estimator based on the estimating equation (2.2) is consistent to $\boldsymbol{\beta}_0$, the true value of $\boldsymbol{\beta}$, and asymptotically normally distributed. When λ and the variance function $\sigma^2(.)$ are unknown, we establish that the consistency and the weak convergence of the $\tilde{\boldsymbol{\beta}}$ still hold with the estimated variance function $\sigma_n^2(.)$ in (2.2). The following regularity conditions are required for the asymptotic results.

We assume the true parameter value $\boldsymbol{\beta}_0$ for $\boldsymbol{\beta}$ belongs to a compact set $\mathcal{B}$ in R^p, and employ the conditions (M4)-(M6), (K1)-(K4) in Chiou and Muller (1999) for the local polynomial regression and conditions (A1) and (A3) in Yu, Liu and Chen (2013) for censored data. The (M4) is a moment assumption of the error term for the uniform consistency of the variance function estimator. The (M5) ensures that the linear predictors are bounded and the (M6) is the assumption for the density function of data mean, which assures that for discrete or binary variables, the number of combinations of levels of the predictor variables is large. The (K1)-(K4) are regularity conditions for kernel function and bandwidth for the uniform consistency of the variance function estimator. The (A1) is a standard assumption of survival analysis for considering only intervals that are bounded from right. The (A3) ensures that the variance of survival time is finite.

Lemma 1. *Under the assumptions (M4)-(M6), (K1)-(K4) and (A1), (A3),* $\tilde{\boldsymbol{\beta}}$ *is a consistent estimator of the parameter* $\boldsymbol{\beta}_0$.

Next, we prove that $\tilde{\sigma}_n^2(.)$ is a uniformly consistent estimator of $\sigma^2(.)$ and show the asymptotic distribution of the $\tilde{\boldsymbol{\beta}}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Lemma 2. *Under the assumptions (M4)-(M6), (K1)-(K4) and (A1), (A3),*

$$sup_{u\in I}|\tilde{\sigma}_n^2(u) - \sigma^2(u)| = O_p\left((\frac{\log n}{nb})^{1/2} + b^2 + \frac{1}{\sqrt{nb}}\right),$$

where $I = support(f_\mu)$ and f_μ is the density of μ, which varies with $\mathbf{X}$.

Theorem 1. *Under the assumptions (M4)-(M6), (K1)-(K4) and (A1), (A3), and the variances $\tilde{\sigma}_n^2(\mu_i)$ are truncated below by a sequence $\zeta_n > 0$, which satisfies the requirements in Chiou and Muller (1999), i.e., $\zeta_n \to 0$, $b/\zeta_n \to 0$, $nb^2\zeta_n^2 \to \infty$, $nb^2\zeta_n^2/logn \to \infty$. Then $\tilde{\boldsymbol{\beta}}$ is asymptotically normally distributed $N(\boldsymbol{\beta}_0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$, where $\mathbf{A} = E(\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X})$, and $\mathbf{B} = E(\mathbf{X}^T\mathbf{W}^{-1}\boldsymbol{\Sigma}\mathbf{W}^{-1}\mathbf{X})$, in which $\mathbf{X}$ is the matrix with rows $\mathbf{X}_i$, $\mathbf{W}$ is a diagonal matrix with elements $\sigma^2(\mu_i)$, $\boldsymbol{\Sigma}$ is the diagonal variance-covariance matrix of $\tilde{y}_i$ with diagonal elements $E(Y_i^2 \mid Y_i < C_i)P(y_i < C_i) - \{E(Y_i \mid Y_i < C_i)P(Y_i < C_i)\}^2 + E(C_i^2 \mid Y_i > C_i, C_i > \mu_i)P(y_i > C_i, C_i > \mu_i) - \{E(C_i \mid Y_i > C_i)P(Y_i > C_i, C_i > \mu_i)\}^2 + \mu^2 P(y_i > C_i, C_i < \mu_i)(1 - P(y_i > C_i, C_i < \mu_i))$.*

Remark: Theorem 1 shows that for censored observations, if the Laplace approximated observation $\tilde{y}_i$ and its variance yield valid approximations to $E(Y_i|Y_i > C_i)$ and $\mathrm{Var}(E(Y_i|Y_i > C_i))$ respectively, then $\tilde{\boldsymbol{\beta}}$ has the same asymptotic distribution as the weighted least squares estimator $\hat{\boldsymbol{\beta}}_W$ (Yu, Liu and Chen (2013)), an estimator based on a discrete estimating function. Specifically, for censoring observations, if $\tilde{y}_i = E(Y_i|Y_i > C_i) + o_p(n^{-1/2})$, and hence $\mathrm{Var}(\tilde{y}_i) = \mathrm{Var}(E(Y_i|Y_i > C_i)) + o_p(n^{-1/2})$, then $\tilde{\boldsymbol{\beta}}$ is consistent to $\boldsymbol{\beta}_0^*$, the true value for $\boldsymbol{\beta}^*$ (Ritov (1990)). For the asymptotic

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

variance, let $\mathrm{Var}(\hat{\boldsymbol{\beta}}_W) = \mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}$, then $\mathbf{B}_w = \mathbf{B}$ based on Theorem 5.1 in Ritov (1990) and

$$\begin{aligned}
\mathbf{A} &= -E(\frac{\partial \tilde{\mathbf{U}}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}) \\
&= -\frac{\partial E(\tilde{\mathbf{U}}_i(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} + \int \tilde{\mathbf{U}}_i(\boldsymbol{\beta}) f'(e_i)\mathbf{X}_i/\sigma(\mu_i) de_i \\
&= \mathrm{Cov}(\tilde{\mathbf{U}}_i(\boldsymbol{\beta}), \frac{\mathbf{X}_i^T}{\sigma(\mu_i)}\frac{f'(e_i)}{f(e_i)}) \\
&= \mathbf{A}_w,
\end{aligned}$$

where $f(.)$ is the density function of $e_i = (\tilde{y}_i - \boldsymbol{\beta}^T\mathbf{X}_i)/\sigma(\mu_i), i = 1, \cdots, n$.

The sketches of the proofs of the Lemmas and Theorem are provided in online supplementary materials.

4 Inference

4.1 The bias of the estimation

Because the $\boldsymbol{\beta}$ satisfying $E\tilde{\mathbf{U}}(\boldsymbol{\beta}) = 0$ may not be the same as the $\boldsymbol{\beta}^*$ in model (1), the estimator by solving $\tilde{\mathbf{U}}(\boldsymbol{\beta}) = 0$ may be a biased estimator of $\boldsymbol{\beta}_0^*$, the true value for $\boldsymbol{\beta}^*$. Next, we will quantify the bias and describe how to adjust.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Because $E(Y_i/\tilde{\sigma}_n(\tilde{\mu}_i)|\mathbf{X}_i) = \boldsymbol{\beta}_0^{*T}\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)$, the bias can be quantified as

$$\begin{aligned}
\mathbf{b}(\boldsymbol{\beta}) &= E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}_0^* \\
&= \boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^* + o_p(1) \\
&= \left(\sum_{i=1}^{n}(\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i))^{\otimes 2}\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)[E(\tilde{Y}_i|\mathbf{X}_i) - E(Y_i|\mathbf{X}_i)]/\tilde{\sigma}_n(\tilde{\mu}_i) + o_p(1) \\
&= \left(\sum_{i=1}^{n}(\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i))^{\otimes 2}\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)[A_i + B_i]/\tilde{\sigma}_n(\tilde{\mu}_i) + o_p(1),
\end{aligned}$$

$$A_i = \int_0^{\mu_i}\int_{C_i}^{\infty}(\mu_i - Y_i)f(Y_i|\mathbf{X}_i)dY_i g_i(C_i)dC_i,$$

and

$$B_i = \int_{\mu_i}^{\infty}\int_{C_i}^{\infty}(C_i - Y_i)f(Y_i|\mathbf{X}_i)dY_i g_i(C_i)dC_i,$$

where $\otimes$ represents Kronecker product, $\tilde{Y}_i$ is a random variable with realizations $\tilde{y}_i$, $g_i(.)$ is the probability density function of the censoring time C_i. If the censoring time is discrete, the integral will be changed into a sum of the censoring values at and above the observed C_i value. It is easy to see that A_i, the bias when $\delta_i = 0, \lambda_i = 0$, can be both positive and negative and B_i, when $\delta_i = 0, \lambda_i = 1$, is always negative.

Because $E(E(Y_i|Y_i > C_i)|\mathbf{X}_i) = E(Y_i|\mathbf{X}_i)$, the bias can be estimated by

$$\hat{\mathbf{b}}(\boldsymbol{\beta}) = \tilde{\boldsymbol{\beta}} - \left(\sum_{i=1}^{n}(\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i))^{\otimes 2}\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)[\delta_i y_i + (1-\delta_i)\hat{E}(Y_i|Y_i > C_i, \mathbf{X}_i)]/\tilde{\sigma}_n(\tilde{\mu}_i),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

where

$$\hat{E}(Y_i|Y_i > C_i, \mathbf{X}_i) = \tilde{\mu}_i^* + \tilde{\sigma}_n(\tilde{\mu}_i)\hat{E}(e_i|e_i > r_i, \mathbf{X}_i),$$

$$\hat{E}(e_i|e_i > r_i, \mathbf{X}_i) = \frac{\int_{r_i}^{\infty} u d\hat{F}(u)}{1 - \hat{F}(r_i)},$$

and

$$\hat{F}(r) = 1 - \prod_{\{j:r_j<r\}} \frac{\delta_j}{\sum_{k=1}^{n} I(r_k \geq r_j)},$$

which is the Kaplan-Meier estimate of the survival function of e_i. Here $e_i = (Y_i - \mu_i^*)/\tilde{\sigma}_n(\tilde{\mu}_i)$ and $r_i = (y_i - \tilde{\mu}_i^*)/\tilde{\sigma}_n(\tilde{\mu}_i)$, $\tilde{\mu}_i^* = \tilde{\boldsymbol{\beta}}^{*T}\mathbf{X}_i$ and

$$\tilde{\boldsymbol{\beta}}^* = \left(\sum_{i=1}^{n}(\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i))^{\otimes 2}\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)\hat{E}(Y_i|Y_i > C_i, \mathbf{X}_i)/\tilde{\sigma}_n(\tilde{\mu}_i).$$

Note that $\hat{E}(Y_i|Y_i > C_i, \mathbf{X}_i)$ includes $\tilde{\boldsymbol{\beta}}^*$, the estimator of $\boldsymbol{\beta}^*$, so we use a iterative algorithm to solve $\tilde{\boldsymbol{\beta}}^*$, which is provided in Section 4.3.

Based on $E(e_i) = 0$ and the work of Ritov (1990), it is easy to see that $\tilde{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0^* = O_p(n^{-1/2})$. As a result, $\hat{\mathbf{b}}(\boldsymbol{\beta})$ is a consistent estimator of $\mathbf{b}(\boldsymbol{\beta})$. Therefore, $\lim_{n\to\infty} p(\boldsymbol{\beta}_0^* \in (\tilde{\boldsymbol{\beta}} - Z_{1-\alpha/2}SE(\tilde{\boldsymbol{\beta}}) - \hat{\mathbf{b}}(\boldsymbol{\beta}), \tilde{\boldsymbol{\beta}} + Z_{1-\alpha/2}SE(\tilde{\boldsymbol{\beta}}) - \hat{\mathbf{b}}(\boldsymbol{\beta}))) = 1 - \alpha$.

4.2 Variance Estimation

Theorem 1 shows that the asymptotic variance of $\tilde{\boldsymbol{\beta}}$ involves the unknown conditional expectation of C_i and Y_i. Although we can estimate the unknown parameters in the asymptotic variance, this may lead to loss of efficiency for small samples. Here we propose to use a bootstrap approach to estimate the variance of $\tilde{\boldsymbol{\beta}}$. Specifically, we can

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

repeatedly sample B times of n observations $\{T_i, \delta_i, \mathbf{X}_i\}$ from the original dataset with replacement, and then use the Laplace approximated weighted least squares method in section 2 to obtain $\{\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_B\}$. The variance of $\tilde{\boldsymbol{\beta}}$ can be estimated by the sample variance of $\tilde{\boldsymbol{\beta}}_j, j = 1, \cdots, B$.

4.3 Algorithm of the proposed method

For practical implementation, we propose iterative updating procedures in the two main parts, calculation of $\tilde{\boldsymbol{\beta}}$ and calculation its bias correction, for the new proposed Laplace approximated weighted least squares method.

- Set the least squares estimator (Jin, Lin and Ying (2006)) of $\boldsymbol{\beta}$ as the initial value $\tilde{\boldsymbol{\beta}}^{(0)}$, and set the initial $\tilde{\sigma}_n^2(.) = 1$.

- At the mth step,

  - Calculate $\tilde{\mu}_i^{(m)} = (\tilde{\boldsymbol{\beta}}^{(m-1)})^T \mathbf{X}_i, \tilde{\lambda}_i^{(m)} = I(y_i > \tilde{\mu}_i^{(m)}), \tilde{y}_i = \delta_i y_i + (1 - \delta_i)\tilde{\lambda}_i^{(m)} y_i + (1 - \delta_i)(1 - \tilde{\lambda}_i^{(m)})\tilde{\mu}_i^{(m)}$.

  - Update variance function $\tilde{\sigma}_n^2(.)$ using the local polynomial regression with the optimal global bandwidth $n^{-1/5}$ and the Epanechnikov Kernel as in Fan and Yao (1998).

  - Obtain $\tilde{\boldsymbol{\beta}}^{(m)}$ based on the weighted least squares method,

$$\tilde{\boldsymbol{\beta}}^{(m)} = \sum_{i=1}^{n} [\mathbf{X}_i^T \mathbf{X}_i / \tilde{\sigma}_n^2(\tilde{\mu}_i^{(m)})]^{-1} \mathbf{X}_i^T \tilde{y}_i / \tilde{\sigma}_n^2(\tilde{\mu}_i^{(m)}).$$

- Stop the iterations when $\tilde{\boldsymbol{\beta}}^{(m)}$ converges and denote the convergent value as $\tilde{\boldsymbol{\beta}}$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

and calculate $\tilde{\mu}$.

- For the bias correction, set the initial values as $\tilde{\boldsymbol{\beta}}^{*(0)} = \tilde{\boldsymbol{\beta}}$ and $\tilde{\mu}_i^{*(0)} = \tilde{\mu}_i$.

- At the lth step,

  - Estimate $\hat{E}^{(l)}(Y_i|Y_i > C_i, \mathbf{X}_i) = \tilde{\mu}_i^{*(l-1)} + \tilde{\sigma}_n(\tilde{\mu}_i)\hat{E}(e_i|e_i > \tilde{r}_i, \mathbf{X}_i)$ as in Section 4.1, where $\tilde{r}_i = (y_i - \tilde{\mu}_i^{*(l-1)})/\tilde{\sigma}_n(\tilde{\mu}_i)$.

  - Calculate the bias by

$$\hat{\mathbf{b}}^{(l)} = \left(\sum_{i=1}^{n}(\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i))^{\otimes 2}\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i/\tilde{\sigma}_n(\tilde{\mu}_i)(\tilde{y}_i - \hat{E}^{(l)}(Y_i|Y_i > C_i, \mathbf{X}_i))/\tilde{\sigma}_n(\tilde{\mu}_i).$$

  - Calculate $\tilde{\boldsymbol{\beta}}^{*(l)} = \tilde{\boldsymbol{\beta}}^{*(l-1)} - \hat{\mathbf{b}}^{(l)}$ and $\tilde{\mu}_i^{*(l)} = (\tilde{\boldsymbol{\beta}}^{*(l)})^T\mathbf{X}_i$.

- Stop the iterations when $\hat{\mathbf{b}}^{(l)}$ converges and denote the bias as $\hat{\mathbf{b}}$ at the convergence.

- Calculate $\tilde{\boldsymbol{\beta}}^* = \tilde{\boldsymbol{\beta}} - \hat{\mathbf{b}}$.

- Calculate the $(1-\alpha)\times 100\%$ confidence interval for $\boldsymbol{\beta}_0^*$ by $(\tilde{\boldsymbol{\beta}}^* - Z_{1-\alpha/2}SE(\tilde{\boldsymbol{\beta}}), \tilde{\boldsymbol{\beta}}^* + Z_{1-\alpha/2}SE(\tilde{\boldsymbol{\beta}}))$, where $SE(\tilde{\boldsymbol{\beta}})$ is the standard deviation of $\tilde{\boldsymbol{\beta}}$ obtained by the bootstrap method in Section 4.2.

5 Simulation Studies

5.1 Simulation settings

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

We considered three scenarios for the simulation studies. The first scenario considered homoscedastic data. The second scenario considered heteroscedastic data with covariate-independent censoring. The third scenario considered heteroscedastic data with covariate-dependent censoring. In the second and third scenarios, we also considered different variance functions.

We report the bias and empirical standard error (SE) to compare the bias and efficiency of the Laplace approximated weighted least squares method (hereafter, we call it Laplace method for convenience) with those of weighted least squares method (WLS) (Yu, Liu and Chen (2013)) and of the local Buckley-James method (LBJ) (Pang, Lu and Wang (2015)) for both homoscedastic and heteroscedastic survival data and of the least squares method (LS) (Buckley and James (1979)) for homoscedastic survival data. In order to check the validity of the variance estimation method proposed for the Laplace method, we report its estimated standard error (SEE) and coverage probability of the 95% confidence interval (Cov) as well. The results were based on 500 Monte Carlo runs for each setting.

5.2 Data simulation

To investigate the performance of the new proposed Laplace method when the variance function depends on different subsets of the covariates, we simulated data with several covariates from the following model as in Pang, Lu and Wang (2015),

$$Y_i \equiv \log(T_i) = \boldsymbol{\beta}^T \mathbf{X}_i + \sigma(\boldsymbol{\beta}^T X_i) e_i,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

where $\mathbf{X}_i = \{1, x_{i1}, x_{i2}, x_{i3}, x_{i4}\}$, and $x_{i1} \sim \text{Unif}(-1, 1)$, $x_{i2} = x_{i1}/3 + 2x_{i5}/3$, $x_{i3} \sim$ Bernoulli(0.5), $x_{i4} \sim$ Bernoulli(0.5), $x_{i5} \sim$ Triangle$(-2, 2)$. The $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4\}$. For scenarios 1 and 2, $\boldsymbol{\beta} = \{1, -1, 2, 1, -1\}$, while $\boldsymbol{\beta} = \{0, -1, 2, 1, -1\}$ for scenario 3. The variance $\sigma_i \equiv \sigma(\boldsymbol{\beta}^T\mathbf{X}_i) = 0.7$ for scenario 1. For scenario 2, we considered two variance functions: $\sigma_{i1} = \exp(-0.5 - \boldsymbol{\beta}^T\mathbf{X}_i)$, which satisfies the variance assumption that variance is a function of the mean in model (2.1); $\sigma_{i2} = \exp(-2.5 - x_1\beta_1 - x_3\beta_3)$, which violates the variance assumption, that is, the variance depends on subsets of the covariates instead of the mean. For scenario 3, we also considered two variance functions: σ_{i1}, which is same as in scenario 2 and $\sigma_{i3} = \exp(-1.5 - x_1\beta_1 - x_2\beta_2 - x_3\beta_3)$, which violates the assumption for variance function. Note that σ_{i3} involves more covariates than σ_{i2}, which means that the domain of σ_{i3} is closer to the mean values than that of σ_{i2}.

Two different error distributions of e_i were considered: standard normal distribution and centered standard extreme value distribution. These two error distributions correspond to two popular distributed survival data, log normal and Weibull survival data. Also, two censoring percentages (CP) were considered to investigate its effect on the estimation. The censoring time C_i was generated from $N(c_1, 2)$ for scenarios 1 and 2, with $c_1 = 3.0$ for 20% censoring and $c_1 = 1.6$ for 40% censoring. For scenario 3, censoring time was generated from $N(c_2, 2)$ when $X_{i3} = 1$ and from $N(c_3, 2)$ when $X_{i3} = 0$, with $c_2 = 1.6, c_3 = 2.2$ for 20% censoring and $c_2 = 0.4, c_3 = 0.8$ for 40% censoring. The sample size n=200 or 400 for each setting. The variance of the proposed

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

estimator was calculated with 500 bootstrap samples.

5.3 Simulation results

The results for scenario 1 are summarized in Table 1. For homoscedastic survival data with covariate independent censoring, all methods are unbiased. For the Laplace method, the bootstrap standard errors matches the empirical standard errors well and the coverages probabilities of the confidence intervals close to the nominal level. The LS method is slightly most efficient than the LBJ method because the homoscedasticity assumption is satisfied in this scenario. The LS and LBJ methods are more efficient than the laplace and WLS methods. One explanation is that the LS and LBJ methods use the ordinary least squares estimating equation, which assumes the data is homoscedastic, while the laplace and WLS methods use weighted least squares estimating equation, which uses the estimated variances.

The results obtained from scenario 2 using normally distributed data are summarized in Table 2. In the case of heteroscedastic survival data with covariate-independent censoring and a variance function σ_1, the LS method exhibits clear bias, especially in situations with heavy censoring (40%). Conversely, the other three methods remain unbiased due to their consideration of the heteroscedasticity nature in the data. The Laplace and WLS methods outperform the LBJ method in terms of efficiency. This is because the Laplace and WLS methods utilize the weighted version of the least squares method and the variance assumption in model (2.1) is satisfied. However, when the variance function σ_2 is used, which fails to satisfy the variance

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Table 1: Simulation results for homoscedastic survival data

distr	n	CP	para	Laplace				WLS		LBJ		LS	
				bias	SE	SEE	Cov	bias	SE	bias	SE	bias	SE
Normal	200	20%	β_1	0.001	0.100	0.125	0.942	-0.005	0.101	-0.011	0.085	-0.010	0.085
			β_2	-0.013	0.106	0.135	0.944	-0.003	0.108	0.004	0.085	0.003	0.084
			β_3	-0.013	0.100	0.131	0.964	-0.009	0.098	-0.004	0.088	-0.004	0.087
			β_4	0.014	0.096	0.132	0.962	0.010	0.096	0.007	0.089	0.007	0.088
	400	20%	β_1	0.014	0.065	0.084	0.948	0.008	0.066	0.001	0.059	0.001	0.059
			β_2	-0.023	0.072	0.093	0.938	-0.012	0.072	0.001	0.058	0.001	0.058
			β_3	-0.007	0.064	0.088	0.952	-0.002	0.064	0.005	0.058	0.005	0.057
			β_4	0.005	0.064	0.088	0.974	-0.001	0.064	-0.007	0.061	-0.007	0.061
	200	40%	β_1	0.002	0.117	0.138	0.910	0.001	0.119	-0.015	0.098	-0.014	0.096
			β_2	-0.027	0.124	0.145	0.928	-0.024	0.127	0.007	0.099	0.005	0.095
			β_3	-0.026	0.118	0.149	0.944	-0.025	0.118	-0.006	0.101	-0.007	0.099
			β_4	0.030	0.113	0.15	0.940	0.028	0.116	0.010	0.100	0.011	0.098
	400	40%	β_1	0.023	0.080	0.097	0.958	0.019	0.080	0.002	0.064	0.002	0.063
			β_2	-0.042	0.088	0.105	0.928	-0.035	0.087	0.001	0.066	0.001	0.064
			β_3	-0.019	0.077	0.103	0.958	-0.015	0.078	0.004	0.065	0.004	0.064
			β_4	0.019	0.075	0.103	0.966	0.014	0.076	-0.004	0.068	-0.004	0.066
Extreme	200	20%	β_1	0.017	0.116	0.154	0.958	0.008	0.116	0.006	0.097	0.005	0.098
			β_2	-0.029	0.131	0.168	0.946	-0.014	0.133	-0.003	0.106	-0.001	0.110
			β_3	-0.011	0.128	0.165	0.964	-0.002	0.125	-0.002	0.120	-0.001	0.121
			β_4	0.019	0.131	0.164	0.960	0.012	0.127	-0.001	0.116	-0.002	0.117
	400	20%	β_1	0.006	0.081	0.105	0.968	-0.004	0.082	-0.009	0.072	-0.010	0.073
			β_2	-0.030	0.093	0.118	0.940	-0.012	0.091	-0.004	0.074	-0.002	0.078
			β_3	-0.012	0.084	0.112	0.952	-0.002	0.084	0.004	0.084	0.005	0.084
			β_4	0.013	0.081	0.112	0.956	0.003	0.081	-0.003	0.074	-0.004	0.075
	200	40%	β_1	0.027	0.139	0.172	0.936	0.023	0.140	0.005	0.114	0.004	0.115
			β_2	-0.056	0.156	0.181	0.920	-0.054	0.161	-0.005	0.121	-0.003	0.128
			β_3	-0.027	0.151	0.188	0.942	-0.028	0.152	-0.011	0.138	-0.009	0.140
			β_4	0.031	0.149	0.187	0.948	0.027	0.151	0.006	0.131	0.005	0.133
	400	40%	β_1	0.025	0.100	0.127	0.956	0.019	0.101	-0.006	0.082	-0.008	0.085
			β_2	-0.059	0.119	0.138	0.924	-0.044	0.118	-0.001	0.085	0.002	0.091
			β_3	-0.026	0.100	0.136	0.964	-0.018	0.100	0.007	0.095	0.008	0.097
			β_4	0.027	0.099	0.136	0.960	0.018	0.098	-0.006	0.087	-0.007	0.086

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Table 2: Simulation results for scenario 2 with normal distribution

n	CP	variance	para	Laplace				WLS		LBJ		LS	
				bias	SE	SEE	Cov	bias	SE	bias	SE	bias	SE
200	20%	σ_1	β_1	-0.007	0.025	0.035	0.970	-0.006	0.024	-0.023	0.121	-0.052	0.133
			β_2	0.013	0.038	0.054	0.964	0.012	0.035	0.016	0.163	0.074	0.187
			β_3	0.005	0.028	0.036	0.966	0.004	0.026	0.003	0.113	0.032	0.121
			β_4	-0.006	0.028	0.037	0.978	-0.005	0.026	0.001	0.121	-0.028	0.130
		σ_2	β_1	-0.001	0.007	0.008	0.928	-0.002	0.011	-0.001	0.007	-0.001	0.007
			β_2	0.000	0.006	0.007	0.944	0.002	0.011	0.000	0.006	0.000	0.006
			β_3	0.000	0.006	0.008	0.940	0.027	0.033	0.000	0.006	0.000	0.006
			β_4	0.001	0.007	0.008	0.928	0.027	0.034	0.001	0.007	0.001	0.007
400	20%	σ_1	β_1	-0.003	0.016	0.025	0.962	-0.001	0.016	-0.010	0.094	-0.040	0.104
			β_2	0.005	0.023	0.041	0.978	0.002	0.023	0.023	0.115	0.083	0.144
			β_3	0.004	0.018	0.026	0.938	0.002	0.018	0.015	0.076	0.046	0.089
			β_4	-0.003	0.017	0.026	0.974	-0.002	0.017	-0.020	0.081	-0.051	0.096
		σ_2	β_1	0.000	0.005	0.006	0.934	-0.002	0.007	0.000	0.005	0.000	0.005
			β_2	0.000	0.004	0.005	0.936	0.002	0.007	0.000	0.004	0.000	0.004
			β_3	0.001	0.004	0.005	0.954	0.017	0.021	0.000	0.004	0.001	0.004
			β_4	0.000	0.004	0.006	0.958	0.015	0.020	0.000	0.004	-0.001	0.004
200	40%	σ_1	β_1	-0.015	0.038	0.048	0.966	-0.011	0.033	-0.036	0.141	-0.106	0.176
			β_2	0.027	0.056	0.072	0.964	0.019	0.048	0.034	0.181	0.174	0.255
			β_3	0.012	0.039	0.050	0.956	0.008	0.035	0.013	0.131	0.082	0.160
			β_4	-0.011	0.038	0.049	0.970	-0.007	0.034	-0.007	0.137	-0.077	0.167
		σ_2	β_1	-0.002	0.008	0.010	0.932	-0.004	0.012	-0.001	0.008	-0.001	0.008
			β_2	0.000	0.007	0.009	0.916	0.003	0.013	0.000	0.007	0.000	0.007
			β_3	0.001	0.007	0.009	0.928	0.027	0.034	0.000	0.007	0.000	0.007
			β_4	0.001	0.008	0.009	0.916	0.026	0.032	0.001	0.008	0.001	0.008
400	40%	σ_1	β_1	-0.006	0.023	0.033	0.956	-0.006	0.022	-0.019	0.108	-0.092	0.141
			β_2	0.012	0.033	0.052	0.976	0.011	0.031	0.039	0.126	0.184	0.219
			β_3	0.007	0.025	0.034	0.956	0.007	0.024	0.026	0.089	0.099	0.129
			β_4	-0.006	0.023	0.034	0.974	-0.005	0.022	-0.028	0.094	-0.102	0.136
		σ_2	β_1	-0.001	0.006	0.007	0.932	-0.002	0.008	-0.001	0.006	-0.001	0.006
			β_2	0.000	0.005	0.006	0.924	0.002	0.008	0.000	0.005	0.000	0.005
			β_3	0.001	0.005	0.006	0.928	0.016	0.020	0.001	0.005	0.001	0.005
			β_4	0.000	0.005	0.007	0.952	0.014	0.020	0.000	0.005	0.000	0.005

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

assumption in model (2.1), the WLS method exhibits the least efficiency, while the Laplace, and LBJ methods demonstrate similar efficiencies. This phenomenon can be explained as follows: when the variance function is misspecified, it has the most significant impact on the WLS method since both the variance function and the synthetic data, which are bundled with the variance function in the estimating equation, are affected. In the case of the Laplace method, the variance function in the estimating equation is affected, but the Laplace approximated data remain unaffected by the misspecified variance function because their values do not directly depend on the variance function. Similarly, for the LBJ method, only the synthetic data is affected by the misspecified variance function because it does not include the variance function in the estimating equation. Additionally, simulations were conducted for scenario 2 using extreme value distributed data, and the results are provided in the online supplementary material. The performance is found to be similar to that observed for data with a normal distribution, indicating the robustness of the results for the AFT model across different error distributions.

The results for scenario 3 are summarized in Table 3 for normally distributed data. This scenario is complex as the data is heteroscedastic and the censoring depends on the covariates, which violates the assumption required by LS method. Clearly, the LS method is biased, while other three methods are unbiased. In cases where the variance functions σ_1 satisfy the assumption in model (2.1), both the Laplace and WLS methods exhibit similar efficiencies, which are superior to that of the LBJ method. This finding

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

aligns with the results observed in scenario 2 with the variance function σ_1. However, for scenarios with σ_3 that violate the variance assumption in model (2.1), the efficiency pattern differs from that observed in scenario 2 with σ_2. In scenario 3 with σ_3, the LBJ method is less efficient than the WLS and Laplace methods. Conversely, in scenario 2 with σ_2, the WLS method is less efficient than the LBJ and Laplace methods. This disparity might arise because σ_3 depends on more covariates than σ_2, causing the domain of σ_3 to be closer to the mean than that of σ_2. Consequently, the violation of the variance function assumption in model (2.1) has a lesser impact on σ_3 compared to σ_2. This suggests that as the domain of the variance function moves further away from the mean, the LBJ method becomes more efficient while the WLS method becomes less efficient. However, the Laplace method consistently remains the most efficient, underscoring its robustness. The simulation results for scenario 3 involving extreme value distributed data can be found in the online supplementary material, and they yield similar outcomes to those obtained for normal distribution.

Overall, the LS method is biased for heteroscedastic survival data. The Laplace method is the most efficient method among the three methods for heteroscedastic survival data. Moreover, it is robust to different variance functions, compared with WLS and LBJ methods. In addition, for the Laplace method, the estimated standard error matches the empirical standard error well and the coverage probability of the confidence interval is close to the nominal level. In terms of computational time on a Lenovo Yoga 12 machine with Processor i5-5300U, the Laplace method is much

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Table 3: Simulation results for scenario 3 with normal distribution

n	CP	variance	para	Laplace				WLS		LBJ		LS	
				bias	SE	SEE	Cov	bias	SE	bias	SE	bias	SE
200	20%	σ_1	β_1	-0.009	0.028	0.037	0.970	-0.006	0.024	-0.024	0.125	-0.056	0.137
			β_2	0.018	0.044	0.059	0.964	0.012	0.036	0.013	0.165	0.078	0.192
			β_3	0.008	0.030	0.039	0.960	0.005	0.026	0.006	0.114	0.050	0.130
			β_4	-0.010	0.031	0.040	0.974	-0.006	0.027	0.003	0.125	-0.029	0.138
		σ_3	β_1	-0.003	0.020	0.026	0.958	-0.003	0.019	-0.009	0.067	-0.020	0.071
			β_2	0.005	0.029	0.038	0.956	0.005	0.029	0.001	0.088	0.023	0.098
			β_3	0.000	0.020	0.027	0.960	0.000	0.020	-0.001	0.059	0.013	0.063
			β_4	0.001	0.016	0.022	0.972	0.001	0.016	0.014	0.068	0.003	0.069
400	20%	σ_1	β_1	-0.002	0.016	0.026	0.964	-0.002	0.016	-0.008	0.097	-0.041	0.107
			β_2	0.004	0.022	0.042	0.978	0.002	0.022	0.021	0.119	0.085	0.151
			β_3	0.003	0.016	0.027	0.968	0.002	0.016	0.015	0.079	0.059	0.098
			β_4	-0.002	0.017	0.027	0.970	-0.002	0.018	-0.020	0.080	-0.053	0.097
		σ_3	β_1	-0.001	0.013	0.017	0.950	-0.001	0.013	0.001	0.050	-0.010	0.054
			β_2	0.002	0.019	0.025	0.972	0.001	0.019	0.000	0.062	0.022	0.071
			β_3	0.001	0.014	0.018	0.954	0.001	0.014	0.001	0.041	0.016	0.046
			β_4	0.001	0.011	0.014	0.966	0.001	0.010	-0.001	0.042	-0.014	0.045
200	40%	σ_1	β_1	-0.012	0.038	0.049	0.968	-0.010	0.033	-0.036	0.143	-0.108	0.181
			β_2	0.024	0.059	0.075	0.958	0.019	0.050	0.032	0.182	0.176	0.259
			β_3	0.011	0.040	0.051	0.958	0.007	0.033	0.017	0.131	0.099	0.169
			β_4	-0.011	0.042	0.052	0.966	-0.008	0.036	-0.005	0.139	-0.074	0.171
		σ_3	β_1	-0.007	0.026	0.035	0.952	-0.005	0.025	-0.015	0.078	-0.039	0.090
			β_2	0.012	0.039	0.051	0.950	0.009	0.037	0.011	0.100	0.059	0.127
			β_3	0.004	0.027	0.036	0.964	0.002	0.026	0.003	0.069	0.030	0.080
			β_4	0.001	0.023	0.031	0.970	0.002	0.022	0.026	0.082	0.002	0.083
400	40%	σ_1	β_1	-0.005	0.023	0.034	0.960	-0.004	0.023	-0.015	0.109	-0.086	0.137
			β_2	0.010	0.033	0.054	0.974	0.006	0.031	0.033	0.128	0.178	0.217
			β_3	0.007	0.023	0.035	0.966	0.005	0.023	0.022	0.090	0.108	0.137
			β_4	-0.007	0.025	0.036	0.960	-0.004	0.025	-0.030	0.092	-0.103	0.135
		σ_3	β_1	-0.003	0.018	0.023	0.938	-0.002	0.017	-0.003	0.059	-0.026	0.067
			β_2	0.007	0.026	0.032	0.960	0.004	0.025	0.006	0.068	0.054	0.093
			β_3	0.005	0.018	0.024	0.962	0.003	0.018	0.004	0.047	0.032	0.059
			β_4	0.001	0.014	0.019	0.980	0.002	0.014	0.009	0.049	-0.018	0.055

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

faster than WLS and LBJ methods. For example, for data with $n = 400$ and 40% censoring, the approximate computational time for Laplace method with 500 bootstrap samples was 30 seconds, while it was 17 minutes for WLS method with 50 bootstrap samples and 3 minutes for LBJ method with 500 resampling procedures. For data with $n = 200$ and 40% censoring, the approximate computational time for Laplace method with 500 bootstrap samples was 10 seconds, while it was 2 minutes for WLS method with 50 bootstrap samples and 1 minute for LBJ method with 500 resampling procedures. Therefore, it is important to note that the sample size has little effect on the computational time of the Laplace method, but has a significant effect on the WLS and LBJ methods as they need iteratively update the Kaplan-Meier estimate, whose computational time relies much on the sample size.

6 Real Data Analysis

First, we applied our proposed method to Stanford Heart Transplant data (Miller and Halpern (1982)). The data comes from the Stanford Heart Transplant program from October 1967 till February 1980. In this time period, 184 patients received heart transplantation. The goal of our analysis is to investigate the effect of age at the time of transplant on the survival time after heart transplant. As in Wei, Ying and Lin (1990), we took 176 observations with survival times at least 10 days. Among them, 107 are uncensored. We used the same model as in Miller and Halpern (1982) and in

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Wei, Ying and Lin (1990),

$$\log_{10} T = \beta_0 + \beta_1 age + \beta_2 age^2 + \epsilon.$$

Yu, Yu and Liu (2009) showed that the data is heteroscedastic. Therefore, we applied the new proposed Laplace method to this dataset. For comparison, we applied the least squares method, local Buckley-James method and weighted least squares method as well.

The results with parameter estimates, estimated standard errors and p-values are shown in Table 4. The p-values were calculated based on the Wald statistic. The least squares method concludes that both age and age^2 are significant. However, this data is heteroscedastic, so the least squares method may be biased based on the simulation studies. We observed that the point estimates of the least squares method are further away from those of other three methods, indicating that the estimation from the least squares method may be biased. As a consequence, the reliability of the Wald test-based p-value is questionable, as it assumes an unbiased estimator. Among the three methods for heteroscedastic survival data, the weighted least squares method is the least efficient and concludes that both age and age^2 are not significant. The Laplace method and the local Buckley-James method reach the same conclusion, i.e., the age is not significant, while age^2 is significant (p-value < 0.05). In addition, these two methods have similar efficiencies. Based on simulation studies, the results indicate that the variance assumption in model (2.1) is not satisfied and the domain of the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Table 4: Results for Stanford Heart Transplant data

	Laplace			WLS			LBJ			LS		
covariate	estimate	SE	p-value	estimate	SE	p-value	estimate	SE	p-value	estimate	SE	p-value
age	0.0541	0.0288	0.060	0.0572	0.0695	0.410	0.0796	0.0413	0.054	0.1072	0.0372	0.004
age^2	-0.0010	0.0004	0.012	-0.0010	0.0009	0.267	-0.0013	0.0005	0.009	-0.0017	0.0005	0.001

variance function is further away from the data mean. Therefore, the weighted least squares method is the least efficient and fails to identify the significance of age^2.

Our second example is the Chemotherapy for Stage B/C colon cancer data in R survival package. This data is from one of the first successful trials of adjuvant chemotherapy for colon cancer. The patients with resected stages B and C colorectal carcinoma were randomly assigned to three treatment groups (just observation, Levamisole treatment and Levamisole plus fluorouracil treatment). Levamisole, a low-toxicity compound, was previously administered to animals to treat worm infestations, while fluorouracil is a moderately toxic chemotherapy agent. The goal of this data is to compare the effects of these different treatments on survival time (in days) by controlling age, sex, differentiation of tumour, time from surgery to registration and perforation of colon. There are a total of 906 observations. Among them, 441 are uncensored. Therefore, we regressed the natural logarithm of the survival time on age (in years), sex (1=male, 0=female), treatment, differentiation of tumour (1=well, 2=moderate, 3=poor), time from surgery to registration (0=short, 1=long) and perforation of colon (1=perforation, 0=no perforation), i.e.,

$$\log(T) = \beta_0+\beta_1 age+\beta_2 sex+\beta_3 rxLev+\beta_4 rxLev5Fu+\beta_5 differ+\beta_6 surg+\beta_7 perfor+\epsilon,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

where *rxLev* is the Levamisole treatment (1=if treatment is Levamisole; 0=other treatments), *rxLev5Fu* is Levamisole+fluorouracil treatment (1=if treatment is Levamisole+fluorouracil, 0=other treatments), *differ* is the differentiation of tumour, *surg* is the time from surgery to registration and *perfor* is perforation of colon.

We applied the new proposed method for this data. However, it takes significant computational time for the least squares method, local Buckley-James method and weighted least squares method due to the large sample size. Therefore, we used the naive subset estimation for the weighted least squares method and the local Buckley-James method. Specifically, we randomly divided the data into 5 subsets, so each of the first 4 subsets has 180 observations and the 5th subset has 186 observations. Because of the small sample size of each subset, we can apply both the weighted least squares method and local Buckley-James method. Then the estimators from each subset were combined to obtain the final estimator. For example, let $\hat{\boldsymbol{\beta}}^{(i)}, i = 1, \cdots, 5$ represent the estimators from ith subset for a method, and the corresponding variances are denoted by $\hat{var}(\hat{\boldsymbol{\beta}}^{(i)}), i = 1, \cdots, 5$. Then the final estimator was calculated by $\sum_{i=1}^{5} \hat{\boldsymbol{\beta}}^{(i)}/5$, and its variance was calculated by $\sum_{i=1}^{5} \hat{var}(\boldsymbol{\beta}^{(i)})/25$. We did not include the results from the least squares method as it is only for homoscedastic survival data and is time-consuming. Instead, we provided the estimation from the Cox proportional hazards model for comparison.

The parameter estimates, estimated standard errors and p-values are summarized in Table 5. The Laplace method and the Cox model reach the same conclusion.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

They showed that the treatment Levamisole+fluorouracil significantly improves the survival time, compared with the observation group (p-value < 0.05). However, the Levamisole treatment alone does not have a significant difference from the observation group in terms of survival time. Among the controlling variables, differentiation of the tumor and time from surgery to registration show a significant effect on survival. The weighted least squares method and the local Buckley-James method are much less efficient than the newly proposed method. This is primarily due to their time-consuming nature when applied to large datasets, requiring the use of subset method that includes less information due to smaller sample sizes. As a result, the significance of the time from surgery to registration is not identified, and the weighted least squares method fails to detect the significance of the Levamisole+fluorouracil treatment as well. These findings indicate that the weighted least squares method and the local Buckley-James method experience decreased efficiency when employed through subset methods for large samples. Consequently, the newly proposed method proves to be significantly more efficient than these two methods.

To compare the computational time, the new proposed method used 3 minutes, the weighted least squares method with subset analysis used 48 minutes and the local Buckley-James method with subset analysis used 30 minutes. To investigate the heteroscedasticity of the data, we plotted the variance function versus the mean based on our proposed method in Figure 1. It clearly shows that the data is heteroscedastic. Therefore, our new proposed method is the most efficient method in the frame of the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Table 5: Results for Chemotherapy for Stage B/C colon cancer data

	Laplace			WLS			LBJ			Cox		
covariate	estimate	SE	p-value	estimate	SE	p-value	estimate	SE	p-value	estimate	SE	p-value
age	-0.003	0.002	0.158	-0.008	0.007	0.268	-0.008	0.005	0.104	0.003	0.004	0.458
sex	0.047	0.079	0.547	0.077	0.142	0.588	0.044	0.119	0.708	-0.028	0.095	0.762
rxLev	-0.018	0.076	0.806	-0.066	0.156	0.673	-0.081	0.143	0.570	-0.010	0.112	0.925
rxLev5Fu	0.225	0.076	0.003	0.258	0.183	0.158	0.324	0.162	0.045	-0.354	0.119	0.003
differ	-0.273	0.063	0.000	-0.566	0.157	0.000	-0.527	0.133	0.000	0.331	0.096	0.001
surg	-0.170	0.068	0.012	-0.291	0.150	0.052	-0.256	0.141	0.071	0.216	0.103	0.037
perfor	-0.027	0.163	0.866	0.024	0.414	0.953	-0.020	0.337	0.950	0.128	0.263	0.626

AFT model to analyze this data.

7 Discussions

We have proposed a Laplace approximated weighted least squares method, currently the most efficient method that can handle large heteroscedastic survival data with conditional independence assumption using the accelerated failure time model. The procedure has two steps: the first step estimates the approximated variance function and the second step performs bias correction.

The procedure has two computational advantages. One is that it avoids the iterative Kaplan-Meier estimation to obtain synthetic data, and the other is that it estimates the variance of $\tilde{\boldsymbol{\beta}}$, instead of $\tilde{\boldsymbol{\beta}}^*$ in the bootstrap procedure. By doing so, it avoids the bias calculation for $\tilde{\boldsymbol{\beta}}$ in each bootstrap procedure and hence saves computational time. The proposed approach is also robust to the variance function because the Laplace approximated survival times are not "bundled" with the variance function, i.e., the Laplace approximated survival times do not depend on the variance function. The simulation results support that the Laplace approximation method performs well for different variance functions.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

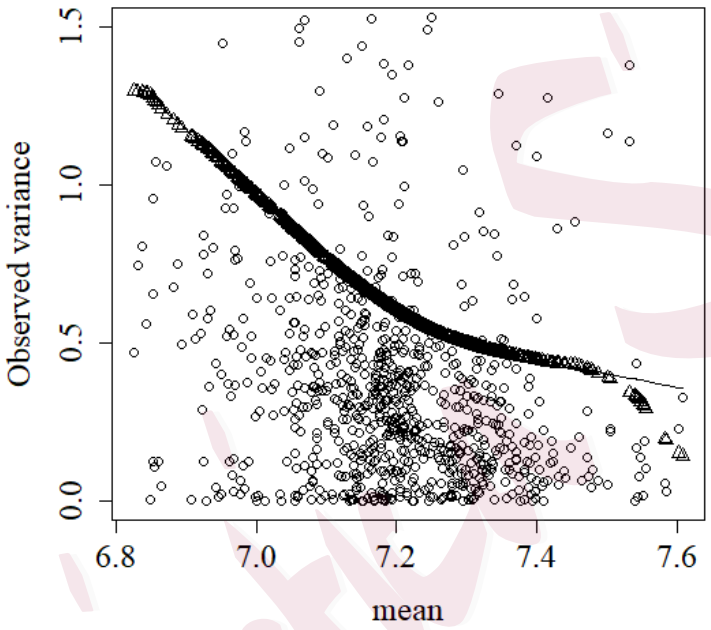

Figure 1: Variance function estimated by the Laplace approximation method for Chemotherapy for Stage B/C colon cancer data.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

The bootstrap procedure proposed in this paper does not include the bias correction step and we justified its validity in Section 4.1. As a result, this approach saves a lot of computational time, especially for large samples. It is also noted that a bias correction step can be added to each of the bootstrap procedures to obtain $\tilde{\boldsymbol{\beta}}^*$. After that, the variance estimation and confidence interval can be obtained via the bootstrap sample $\tilde{\boldsymbol{\beta}}_1^*, \cdots, \tilde{\boldsymbol{\beta}}_B^*$. Our additional simulation studies (not shown here) show that the coverages of the confidence intervals are similar as those obtained from bootstrap sample $\tilde{\boldsymbol{\beta}}_1, \cdots, \tilde{\boldsymbol{\beta}}_B$.

The conditional independence assumption used in this paper is equivalent to the independent censoring in Andersen et al. (1993) for the heteroscedastic AFT model (2.1). Kleinbaum and Klein (2012) interpret this assumption as the probability of being censored for any subject in the risk set with the same covariate values at time t does not depend on that subject's prognosis for failure at time t. This assumption makes sense in many practical applications, but may not hold for some situations. For example, a subject with severe situations at time t may have larger probability to drop out the study (censored at time t). Researchers then proposed to relax this assumption and developed new methodologies for survival data with dependent censoring: see Emura and Chen (2018), Deresa and Van Keilegom (2020), Czado and Van Keilegom (2023), to name a few. Future research can extend our new proposal to survival data with dependent censoring.

The Laplace approximation method appears quite general and yields continuous

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

and differentiable estimating equations, and hence it facilitates both computational and theoretical development. It can be employed to more complex settings, such as heteroscedastic survival data with cure or frailty. This research will advance the applications of the AFT model to the analysis of various survival data.

Supplementary Material

The proofs of the asymptotic properties in Section 3, and additional simulation results in Section 5 are available in supplementary material.

References

Andersen, PK., Borgan, O., Gill, RD. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Breslow, N. E., and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.

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

Butler, R. W., and Wood, A. T. A. (2002). Laplace approximations for hypergeometric functions with matrix argument. *Annals of Statistics* **30**, 1155-1177.

Carroll, R. (1982). Adapting for Heteroscedasticity in Linear Models. *Annals of Statistics* **10**, 1224-1233.

Chen S. and Khan S. (2000). Estimating censored regression models in the presence

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

of nonparametric multiplicative heteroscedasticity. *Journal of Econometrics* **98**, 283-316.

Chiou, J. and Muller H. (1999). Nonparametric Quasi-likelihood. *The Annals of Statistics* **27**, 36-64.

Clark, TG., Bradburn, MJ., Love, SB and Altman, DG. (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer* **89**, 232-238.

Cox, DR. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B* **34**, 187-202.

Czado, C and Van Keilegom, I. (2023). Dependent censoring based on parametric copulas. *Biometrika*, **110**, 721–738.

Deresa, NW. and Van Keilegom, I. (2020). On semiparametric modelling, estimation and inference for survival data subject to dependent censoring. *Biometrika* **107**, 1-16.

Ding, Y., and Nan, B. (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Annals of Statistics* **39**, 3032-3061.

Emura, T. and Chen, Y. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*. Springer, New York.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Fan, J. and Gijbels, I. (1994). Censored regression: Local linear approximations and their applications. *Journal of the American Statistical Association* **89**, 560-570.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.

Fan, J. and Yao, Q. (1998). Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika* **85**, 645-660.

Harding, M. C., and Hausman, J. (2011). Using a laplace approximation to estimate the random coefficients logit model by nonlinear least squares. *International Economic Review* **48**, 1311-1328.

Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics* **59**, 273-297.

Heyde, C. (1997). *Quasi-Likelihood And Its Application*. Springer, New York.

Jin, Z. Z., Lin, D. Y., Wei, L. J. and Ying, Z. L. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Jin, Z. Z., Lin, D. Y. and Ying, Z. L. (2006). On least-squares regression with censored data. *Biometrika* **90**, 147–161.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. New York: Wiley.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Kleinbaum, DG. and Klein, M. (2012). *Survival Analysis: A Self-Learning Text.* Springer: New York.

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

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

Lai, T. L. and Ying, Z. L. (1991b). Rank regression methods for left-truncated and right-censored data. *Annals of Statistics* **19**, 531–556.

Laplace, PS. (1774). Memoir on the probability of causes of events. *Statistical Science* **1**, 366–367.

Leurgans, S. (1987). Linear models, random censoring, and synthetic data. *Biometrika* **74**, 301-309.

Lin, D. Y. and Ying, Z. L. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *Journal of Statistical Planning and Inference* **44**, 47–63.

Liu, W. and Lu, X. (2009). Weighted least squares method for censored linear models. *Journal of Nonparametric Statistics* **21**, 787-799.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

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

Pang, L., Lu, W. and Wang, H. (2015). Local Buckley-James estimation for heteroscedastic accelerated failure time model. *Statistica Sinica* **25**, 863-877.

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

Robins, J. and Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* **79**, 311–319.

Stare, J., Heinzl, H. and Harrell, F. (2000). On the use of Buckley and James least squares regression for survival data. *New Approaches in Applied Statistics* **16**, 125-134.

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

Wedderburn, R.W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-447.

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

Ying, Z. L. (1993). A large-sample study of rank estimation for censored regression data. *Annals of Statistics* **21**, 76–99.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Yu, L. (2011). Nonparametric quasi-likelihood for right censored data. *Lifetime data analysis* **17**, 594-607.

Yu, L., Liu, L.and Chen, D. (2013). Weighted least-squares method for right-censored data in accelerated failure time model. *Biometrics* **69**, 358-365.

Yu, L. and Peace, E. K. (2012). Spline nonparametric quasi-likelihood regression within the frame of the accelerated failure time model. *Computational Statistics and Data Analysis* **56**, 2675-2687.

Yu, L., Yu, R. and Liu, L. (2009). Quasi-likelihood for right-censored data in the generalized linear model. *Communications in Statistics - Theory and Method* **38**, 2187-2200.

Yu, L., Yu, R., Liu, L. and Chen, D. (2012). Extended quasi-likelihood in the generalized linear model for right-censored Data. *Statistics in Medicine* **31**, 1369-1379.

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association* **102**, 1387–1396.

Zhang, M. and Davidian, M. (2008). "Smooth" semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics* **64**, 567-576.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Zhou, M., Bathke, A. and Kim, M. (2012). Empirical likelihood analysis for the heteroscedastic accelerated failure time model. *Statistica Sinica* **22**, 295-316.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0229

Georgia Southern University

E-mail: (lyu@georgiasouthern.edu)

phone: 9124781278

Columbia University

E-mail: (zj7@cumc.columbia.edu)