Extending Buckley–James method for heteroscedastic survival data

Lili Yu[a,*], Liang Liu[b] and Ding-Geng(Din) Chen[c,d]

[a] Department of Biostatistics, Epidemiology and Environmental Health Sciences, JPH College of Public Health, Georgia Southern University, Statesboro, GA, USA;

[b] Department of Statistics, University of Georgia, Athens, GA, USA;

[c] College of Health Solutions, Arizona State University, Phoenix, AZ, USA;

[d] Department of Statistics, University of Pretoria, Pretoria, South Africa

*CONTACT Lili Yu. Department of Biostatistics, Epidemiology and Environmental Health Sciences, JPH College of Public Health, Georgia Southern University, Statesboro, GA, USA. Email: lyu@georgiasouthern.edu

Abstract

The Buckley–James method for the classical accelerated failure time model has been extended to accommodate heteroscedastic survival data in two ways. The first is the weighted least squares method [Yu et al. Weighted least-squares method for right-censored data in accelerated failure time model. Biometrics. 2013;69:358–365], which estimates the heteroscedasticity nonparametrically, while the second is the local Buckley–James method [Pang et al. Local Buckley–James estimation for heteroscedastic accelerated failure time model. Stat Sin. 2015;25:863–877], which uses local Kaplan–Meier method to estimate the heteroscedasticity. However, no comparisons have been done for these two methods. Furthermore, there is no hypothesis testing procedure for this heteroscedastic accelerated failure time model. This paper is then aimed to fill these two gaps to compare the two methods theoretically and numerically with extensive simulation studies. In addition, we propose a class of hypothesis tests for the parameters to provide a complete procedure for analysing heteroscedastic survival data. Two real data examples are used for practical illustration of the comparison and the new proposed tests.
.

Keywords: Accelerated failure time model; local Buckley–James method; weighted least squares method; survival analysis

1. Introduction

The classical accelerated failure time (AFT) model [1,2] provides an attractive alternative to the Cox model [3] due to its direct physical interpretation. Many analysis methods have been proposed for the AFT model. Among them, the rank method [4–10], Buckley–James method [11–14] and profile likelihood method [15] are the main methods.

The classical AFT model only considers the homoscedastic variance of the survival data. However, the heteroscedastic survival data, in which the variance of the survival data depends on the covariates, are often encountered in real applications. Stare et al. [16] showed that the Buckley–James estimator is biased for heteroscedastic survival data. For remedy, some studies have been done to develop different methodologies for heteroscedastic AFT model. Zhang and Davidian [17] proposed a parametric method for heteroscedastic survival data. Chen and Khan [18] investigated a method when the censoring variables are fixed and observed. Zhou et al. [19] used empirical likelihood approaches to handle heteroscedastic survival data. Heuchenne and Van Keilegom [20] proposed a method for a polynomial regression with only one covariate by replacing the censored observations with synthetic data, which are constructed nonparametrically using kernel smoothing method. Liu and Lu [21] proposed a weighted least squares method, in which the censored observations are replaced by synthetic data using inverse probability weighting method [22], and the heteroscedasticity is handled by kernel smoothing.

One of the main inference methods for the classical AFT model, the Buckley–James method, in which the synthetic data are estimated based on Kaplan–Meier estimate, has been extended to handle heteroscedastic survival data in two ways. First, Yu [23], Yu et al. [24], Yu and Peace [25], Yu et al. [26] proposed weighted least squares methods by extending the Buckley–James method and estimating the heteroscedasticity nonparametrically using different methods. Then the method [26] has been extended for testing the homoscedasticity in the AFT model [27]. And second, Pang et al. [28] developed a local Buckley–James method by extending the Buckley–James method using a local Kaplan–Meier method to construct the synthetic data and accommodate the heteroscedasticity simultaneously. They showed through simulations that their method with synthetic data obtained by Kaplan–Meier method is more efficient than that based on inverse probability weighting principle [21]. Later, it has been employed in two-stage estimation of structural instrumental variable models [29]. However, the performance of the methods based on the Kaplan–Meier estimate [26,28] was not systematically compared. Furthermore, there is no hypothesis testing procedure developed for this heteroscedastic AFT model in association with both methods.

This paper is then aimed to fill these two knowledge gaps with a systematic investigation both theoretically and numerically. We will first compare the two methods [26,28] by extending the Buckley–James method theoretically and numerically. Then we develop a class of hypothesis tests for any subsets of parameters in the heteroscedastic AFT model and evaluate their performance through simulations. These investigations would facilitate practitioners to choose the most appropriate analysis method as well conduct complete statistical inference for real heteroscedastic survival data.

This paper is then organized as follows. Section 2 introduces the data and heteroscedastic AFT model. Section 3 describes the weighted least squares method [26] and the local Buckley–James method [28]. Section 4 theoretically compares the weighted least squares method and the local Buckley–James method asymptotically in a proven theorem. In addition, a class of asymptotic hypothesis tests is developed for the parameters associated with heteroscedastic AFT model. Simulation studies for the comparison of these two methods and the evaluations of the proposed hypothesis test are summarized in Section 5 followed by two real data analyses in Section 6. We conclude with some discussions in Section 7. The proofs of the theoretical comparison of the two methods are provided in Appendix.

2. Survival data and the associated heteroscedastic AFT model

Let T_i be the survival time and C_i be the log of censoring time with $Y_i = \log(T_i)$. The observations are denoted as $(y_i, \delta_i), i = 1, \ldots, n,$ where $y_i = min(Y_i, C_i)$ is the observed log of survival time and $\delta_i = I(Y_i \le C_i)$ is the censoring indicator. The heteroscedastic AFT model [26] can then be written as follows:

$Y_i = \mu_i + \epsilon_i = \beta^T \mathbf{X}_i + \sigma\left(\mu_i\right) e_i, \quad i = 1, \ldots, n,$ (1) where β is a *p*-vector of parameters, $\mathbf{X}_i$ is a *p*-vector of covariates with one as the first element, $\sigma(\mu_i)$ is the standard deviation of Y_i and e_i is the random error with unspecified distribution function $F(.)$ with mean 0 and variance 1. The C_i is assumed to be independent of Y_i given $\mathbf{X}_i$.

3. The two methods

3.1. The weighted least squares method

3.1.1. The methodology

The weighted least squares method estimates the parameter β using the estimating equation

$\mathbf{U}\left(\beta, \sigma_n^2(\boldsymbol{\mu})\right) = \sum_{i=1}^{n} \frac{1}{\sigma_n^2(\mu_i)} \mathbf{X}_i \left(\tilde{y}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right) - \beta^T \mathbf{X}_i\right)$, where $\boldsymbol{\mu}$ is a vector of $\mu_i, i = 1, \ldots, n,$ and $\sigma_n^2(\mu_i)$ is a nonparametric estimator of $\sigma^2(\mu_i)$. In addition,

$\tilde{y}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right) = y_i\delta_i + \left(1 - \delta_i\right)\left(\mu_i + \sigma_n(\mu_i)\widetilde{E}\left(e_i \middle| e_i > \tilde{e}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right)\right)\right)$, where

$\widetilde{E}\left(e_i \middle| e_i > \tilde{e}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right)\right) = \int_{\tilde{e}_i(\beta, \sigma_n^2(\mu_i))}^{\infty} \frac{t d\widetilde{F}\left(t\right)}{1 - \widetilde{F}\left(\tilde{e}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right)\right)}$ with

$\tilde{e}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right) = \left(y_i - \beta^T \mathbf{X}_i\right) / \sigma_n\left(\mu_i\right)$ and

$\widetilde{F}\left(e\right) = 1 - \pi_{\{i: \tilde{e}_i(\beta, \sigma_n^2(\mu_i)) \le e\}}\left(1 - \frac{\delta_i}{\sum_{j=1}^{n} I\left(\tilde{e}_j\left(\beta, \sigma_n^2\left(\mu_j\right)\right) \ge \tilde{e}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right)\right)}\right)$. The weighted least squares estimator of β_0, the true value of β, is $\tilde{\beta}$, which is a zero-crossing of the estimating equation $\mathbf{U}(\beta, \sigma_n^2(\boldsymbol{\mu})) = 0.$ In Yu et al. [26], the nonparametric estimator $\sigma_n^2(.)$ is approximated by local polynomial regression given by

$$\sigma_n^2\left(u\right) = \sum_{i=1}^{n} W_{ni}\left(u\right) \tilde{\epsilon}_i^2\left(\beta, \sigma_n^2\left(\mu_i\right)\right), \quad \text{for } i = 1, \ldots, n, \text{ where}$$

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

$A_{n,j}\left(u\right) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{\mu_i - u}{b}\right)\left(\mu_i - u\right)^j, j = 0, 1, 2$ [30];

$\tilde{\epsilon}_i^2\left(\beta, \sigma_n^2\left(\mu_i\right)\right) = \left(\tilde{y}_i\left(\beta, \sigma_n^2\left(\mu_i\right)\right) - \beta^T \mathbf{X}_i\right)^2$ are the 'observed' variance values; *b* is the bandwidth and *K* is the kernel function. In the iterative procedure of the weighted least squares method, the above variance function $\sigma_n^2(u)$ is estimated by

$\tilde{\sigma}_n^2\left(u\right) = \sum_{i=1}^{n} \widetilde{W}_{ni}\left(u\right) \tilde{\epsilon}_i^2\left(\tilde{\beta}, \tilde{\sigma}_n^2\left(\widetilde{\mu}_i\right)\right), \quad i = 1, \ldots, n,$ (2) where $\widetilde{W}_{ni}\left(u\right)$ and $\widetilde{\mu}_i$ are $W_{ni}(u)$ and μ_i evaluated at the weighted least squares estimator $\tilde{\beta}$.

3.1.2. The estimation algorithm

An iterative algorithm was developed for parameter estimation as follows:

- Initialize $\tilde{\beta}$ with the Buckley–James estimator $\tilde{\beta}^{(0)}$ [14]. The initial estimator for $\sigma_n^2(\mu_i), i = 1, \ldots, n$, is $\tilde{\sigma}_n^{2(0)}\left(\widetilde{\mu}_i\right) = 1$ and $\widetilde{\mu}_i = \tilde{\beta}^{(0)T}\mathbf{X}_i$.
- At the mth step,
    - calculate $\tilde{y}_i\left(\tilde{\beta}^{(m)}, \tilde{\sigma}_n^{2(m)}\left(\widetilde{\mu}_i\right)\right) = y_i\delta_i + \left(1-\delta_i\right)\left(\widetilde{\mu}_i + \tilde{\sigma}_n^{(m)}\left(\widetilde{\mu}_i\right)\right.$ $\left.\widetilde{E}\left(e_i \middle| e_i > \tilde{e}_i\left(\tilde{\beta}^{(m)}, \tilde{\sigma}_n^{2(m)}\left(\widetilde{\mu}_i\right)\right)\right)\right)$, where $\widetilde{\mu}_i = \tilde{\beta}^{(m)T}\mathbf{X}_i$.
    - Update the variance function estimator $\tilde{\sigma}_n^{2(m+1)}\left(\widetilde{\mu}_i\right)$ by local kernel smoothing method by (2).
    - Update the parameter estimators by the weighted least squares method: $\tilde{\beta}^{(m+1)} = \sum_{i=1}^{n}\left(\mathbf{X}_i^T\tilde{\sigma}_n^{-2(m+1)}\left(\widetilde{\mu}_i\right)\mathbf{X}_i\right)^{-1}\mathbf{X}_i^T\tilde{\sigma}_n^{-2(m+1)}\left(\widetilde{\mu}_i\right)\tilde{y}_i\left(\tilde{\beta}^{(m)T}, \tilde{\sigma}_n^{2(m)}\left(\widetilde{\mu}_i\right)\right)$.
- The iteration is completed until $\left|\tilde{\beta}^{(m+1)} - \tilde{\beta}^{(k)}\right| < d$, where $k \in \{0, 1, \ldots, m\}$ and d is the prespecified convergence criterion. This stopping rule considers possible oscillation among iterations due to the discrete estimating function.

The variance of the weighted least squares estimator is estimated by the regular bootstrap method, which described in Yu et al. [26] in detail.

3.2. The local Buckley–James method

3.2.1. The methodology

The local Buckley–James method [28] estimates β using the following estimating equation: $\mathbf{L}\left(\beta\right) = \sum_{i=1}^{n}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\hat{y}_i\left(\beta\right) - \beta^T\mathbf{X}_i\right)$, where $\overline{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$ and $\hat{y}_i\left(\beta\right) = \delta_i y_i + \left(1-\delta_i\right)\left(\beta^T\mathbf{X}_i + \widehat{E}\left(\epsilon_i \middle| Y_i > C_i, \beta^T\mathbf{X}_i\right)\right)$. Further $\widehat{E}\left(\epsilon_i \middle| Y_i > C_i, \beta^T\mathbf{X}_i\right) = \frac{\int_{\hat{\epsilon}_i(\beta)}^{\infty} u d\widehat{F}^{\epsilon}\left(u \middle| \beta^T\mathbf{X}_i\right)}{1-\widehat{F}^{\epsilon}\left(\hat{\epsilon}_i\left(\beta\right) \middle| \beta^T\mathbf{X}_i\right)}$, where $\hat{\epsilon}_i\left(\beta\right) = y_i - \beta^T\mathbf{X}_i$ and $\widehat{F}^{\epsilon}\left(t \middle| \beta^T\mathbf{X}_i\right) = 1 - \prod_{j:\hat{\epsilon}_j(\beta)<t}\left\{1 - \frac{B_{nj}\left(\beta^T\mathbf{X}_i\right)\delta_j}{\sum_{k=1}^{n} I\left(\hat{\epsilon}_k\left(\beta\right) \geq \hat{\epsilon}_j\left(\beta\right)\right)B_{nk}\left(\beta^T\mathbf{X}_i\right)}\right\}$ is the local Kaplan–Meier estimator of $F^{\epsilon}\left(t \middle| \beta^T\mathbf{X}_i\right) = p\left(\epsilon_i \leq t \middle| \beta^T\mathbf{X}_i\right)$ in which $B_{nk}, k = 1, \ldots, n$, is a sequence of non-negative weights with $\sum_{k=1}^{n} B_{nk} = 1$ and Nadaraya–Watson type of weight [31] is used, such that $B_{nk}\left(\beta^T\mathbf{X}_i\right) = \frac{K\left(\frac{\beta^T\mathbf{X}_i - \beta^T\mathbf{X}_k}{b}\right)}{\sum_{l=1}^{n} K\left(\frac{\beta^T\mathbf{X}_i - \beta^T\mathbf{X}_l}{b}\right)}$, in which b is the bandwidth such that $b \to 0$ as $n \to \infty$ and K is a symmetric kernel function. The local Buckley–James estimator $\hat{\beta}$ is obtained by solving $\mathbf{L}(\beta) = 0$.

3.2.2. Algorithm of local Kaplan–Meier estimation

The following iterative procedure is used to obtain the estimator $\ddot{\beta}$ for local Buckley–James method.

- Initialize $\hat{\beta}$ with the Buckley–James estimator $\hat{\beta}^{(0)}$.
- At the *m*th step,
  - Calculate $\hat{y}_i\left(\hat{\beta}^{(m)}\right) = \delta_i y_i + \left(1-\delta_i\right)\left(\hat{\beta}^{(m)T}\mathbf{X}_i + \widehat{E}\left(\epsilon_i \middle| Y_i > C_i, \hat{\beta}^{(m)T}\mathbf{X}_i\right)\right)$.
  - Update the parameter estimators by least squares method, $\hat{\beta}^{(m+1)} = \sum_{i=1}^{n}\left(\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\right)^{-1}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T \hat{y}_i\left(\hat{\beta}^{(m)}\right)$.
- The iteration is completed until convergence criterion is achieved.

3.2.3. The resampling method

The resampling method was used to estimate the variances of the local Buckley–James estimators. It introduces the random perturbation to the estimating equation as $\mathbf{L}^*\left(\beta\right) = \sum_{i=1}^{n} a_i\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\hat{y}_i^*\left(\beta\right) - \beta^T\mathbf{X}_i\right)$, where a_i is the perturbation variable generated from the standard exponential distribution, $\hat{y}_i^*\left(\beta\right) = \delta_i y_i + \left(1-\delta_i\right)\left(\beta^T\mathbf{X}_i + \widehat{E}^*\left(\epsilon_i \middle| Y_i > C_i, \beta^T\mathbf{X}_i\right)\right)$, $\widehat{E}^*\left(\epsilon_i \middle| Y_i > C_i, \beta^T\mathbf{X}_i\right) = \frac{\int_{\hat{\epsilon}_i(\beta)}^{\infty} u d\widehat{F}^{\epsilon*}\left(u \middle| \beta^T\mathbf{X}_i\right)}{1-\widehat{F}^{\epsilon*}\left(\hat{\epsilon}_i\left(\beta\right) \middle| \beta^T\mathbf{X}_i\right)}$ and $\widehat{F}^{\epsilon*}\left(t \middle| \beta^T\mathbf{X}_i\right) = 1 - \prod_{j:\hat{\epsilon}_j(\beta)<t}\left\{1 - \frac{B_{nj}\left(\beta^T\mathbf{X}_i\right)\delta_j a_j}{\sum_{k=1}^{n} a_k I\left(\hat{\epsilon}_k\left(\beta\right) \geq \hat{\epsilon}_j\left(\beta\right)\right)B_{nk}\left(\beta^T\mathbf{X}_i\right)}\right\}$.

Next, we use the estimating equation $\mathbf{L}^*(\beta)$ to obtain the estimator $\hat{\beta}^*$ by the algorithm in Section 3.2.2 with the initial estimate $\hat{\beta}$. The resampling procedure is repeated *N* times to produce $\hat{\beta}_l^*, l = 1, \ldots, N$. The variance of $\hat{\beta}$ is estimated by the sample variance of $\hat{\beta}_l^*, l = 1, \ldots, N$, due to the fact that, given the observed data, $\sqrt{n}\left(\hat{\beta}^* - \hat{\beta}\right)$ has the same limiting distribution as $\sqrt{n}\left(\hat{\beta} - \beta_0\right)$.

4. Asymptotic property of the weight least squares estimator and the local Buckley–James estimator

4.1. Theoretical comparison

With the above development, we can now compare the weighted least squares estimator [26] and the local Buckley–James estimator [28] theoretically.

Since the weighted least squares method is a weighted version of the ordinary least squares method and the local Buckley–James method takes the form of the ordinary least square method, this would indicate that the weighted least squares estimator could be more efficient than the local Buckley–James estimator when the data is heteroscedastic, while they may be the same when data is homoscedastic, although both are designed for heteroscedastic data. This property is stated and proved by the following theorem (detailed proof in Appendix).

Theorem 4.1

Under the common assumptions for the weighted least squares method [26] and the local Buckley–James method [28], the weighted least squares estimator [26] is asymptotically more efficient than the local Buckley–James estimator [28] when the data is heteroscedastic and they are the same asymptotically when the data is homoscedastic.

4.2. Hypothesis test

With the estimated parameters from the heteroscedastic AFT model, the next logical step is to develop statistical significance tests involving several components of the parameters. For this purpose, we propose a class of hypothesis tests based on the weighted least squares method [26] and the local Buckley–James method [28] as follows:
$H_0 : \mathbf{\Lambda}\beta = \boldsymbol{\eta}_0 \quad \textbf{versus} \quad H_1 : \mathbf{\Lambda}\beta = \boldsymbol{\eta}_1,$ where $\mathbf{\Lambda}$ is an $m \times p$ matrix with rank *m*, $m < p$ and $\boldsymbol{\eta}_0$, $\boldsymbol{\eta}_1$ are $m \times 1$ vectors. The tests for individual parameters are special cases for the proposed test.

We consider the two test statistics, G_1 based on the weighted least squares method and G_2 based on the local Buckley–James method, as follows:
$G_1 = \left(\mathbf{\Lambda}\tilde{\beta} - \boldsymbol{\eta}_0\right)^T \left(\mathbf{\Lambda}\widetilde{\mathbf{\Sigma}}^{-1}\mathbf{\Lambda}^T\right)^{-1} \left(\mathbf{\Lambda}\tilde{\beta} - \boldsymbol{\eta}_0\right),$ where $\widetilde{\mathbf{\Sigma}}$ is the estimated variance–covariance matrix for $\tilde{\beta}$, which is obtained by the sample variance–covariance matrix of the bootstrap samples in Section 3.1.2, and
$G_2 = \left(\mathbf{\Lambda}\hat{\beta} - \boldsymbol{\eta}_0\right)^T \left(\mathbf{\Lambda}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Lambda}^T\right)^{-1} \left(\mathbf{\Lambda}\hat{\beta} - \boldsymbol{\eta}_0\right),$ where $\widehat{\mathbf{\Sigma}}$ is the estimated variance–covariance matrix for $\hat{\beta}$, which is obtained by the sample variance–covariance matrix of the resampling samples described in Section 3.2.3.

Under H_0, it is easy to see that both G_1 and G_2 follow χ^2_m, i.e. χ^2 distribution with *m* degrees of freedom, while they follow $\chi^2_m(\rho^2)$, i.e. a non-central χ^2 distribution with *m* degrees of freedom and the non-centrality parameter ρ^2, where ρ^2 is a real constant satisfying
$\left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0\right)^T \left(\mathbf{\Lambda}\widetilde{\mathbf{\Sigma}}^{-1}\mathbf{\Lambda}^T\right)^{-1} \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0\right) \to \rho^2$ in G_1 and $\left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0\right)^T \left(\mathbf{\Lambda}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Lambda}^T\right)^{-1} \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0\right) \to \rho^2$ in G_2. Then for the first test, the null hypothesis is rejected if $G_1 > \chi^2_{m,\alpha}$ for a level of α test, where $\chi^2_{m,\alpha}$ is the $100(1-\alpha)$ quantile of χ^2_m. Similarly, it is rejected if $G_2 > \chi^2_{m,\alpha}$ for the second test.

5. Simulation studies

We conduct simulation studies to compare the weighted least squares method (WLS) and the local Buckley–James method (LBJ) in finite samples and further evaluate the performance of the proposed hypothesis test. The simulation codes are available from the first author upon request.

5.1. Simulation design

There are totally four scenarios designed for the simulation studies. Scenario 1 considers survival data with homoscedastic errors and covariate-independent censoring. Scenario 2 considers survival data with heteroscedastic errors and covariate-independent censoring. Scenario 3 considers survival data with heteroscedastic error and covariate-independent censoring with more covariates. Scenario 4 considers survival data with heteroscedastic errors and covariate-dependent censoring.

5.2. Data generation

The survival data are generated from the following model:
$Y_i = \beta^T \mathbf{X}_i + \sigma\left(\beta^T \mathbf{X}_i\right) e_i$, where $\mathbf{X}_i = \{1, X_{i1}, X_{i2}\}$ for scenarios 1, 2 and 4, where $X_{i1} \sim Unif(-1,1), X_{i2} \sim Bernoulli(0.5)$ and $\mathbf{X}_i = \{1, X_{i1}, X_{i2}, X_{i3}, X_{i4}\}$ for scenario 3, where $X_{i1} \sim Unif(-1,1)$, $X_{i2} = X_{i1}/3 + 2X_{i5}/3$; $X_{i5} \sim Triangle(-2,2)$, which represents triangle distribution with left triangle end point -2 and right triangle end point 2; $X_{i3} \sim Bernoulli(0.5)$ and $X_{i4} \sim Bernoulli(0.5)$. Note that the covariates X_1 and X_2 are correlated. The parameter vector $\beta = (0,1,1)^T$ for scenarios 1, 2, 4 and $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T = (6,-1,2,1,-1)^T$ for scenario 3. The variance $\sigma_i = \sigma\left(\beta^T \mathbf{X}_i\right) = 0.7$ for scenario 1, $\sigma_i = exp\left(-0.3 - \beta^T \mathbf{X}_i\right)$ for scenarios 2 and 4. Because there are more covariates for scenario 3, we consider the effect of different variance functions on the estimation. We consider three variance functions: $\sigma_{i1} = exp\left(3.52 - \beta^T \mathbf{X}_i\right)$, which satisfies the assumption that variance is a function of the mean in model (1);
$\sigma_{i2} = exp(-3.52 - x_1\beta_1 - x_2\beta_2 - x_3\beta_3)$, which violates this assumption; $\sigma_{i3} = exp(-3.52 - x_4\beta_4)$, which violates this assumption more than σ_{i2} because it less depends on the mean values.

Furthermore, two families of error distribution for e_i are considered with standard normal distribution and centred standard extreme value distribution for scenarios 1, 2, 4, but only standard normal distribution for scenario 3. Two censoring percentages (CP) are considered for each scenario. The censoring time C_i is generated from $N(c_1, 2)$ for scenarios 1, 2, 4, where $c_1 = 2.4$ for 20% censoring and $c_1 = 1.1$ for 40% censoring. For scenario 3, the censoring time is generated from $N(c_2, 2)$ when $X_{i2} = 1$ and from $N(c_3, 2)$ when $X_{i2} = 0$ with $c_2 = 1.6, c_3 = 2.9$ for 20% censoring and $c_2 = 0.6, c_3 = 1.9$ for 40% censoring.

We consider two sample sizes for each setting, 200 and 400. The variance of the WLS estimator was calculated by bootstrap method with 50 bootstrap samples [26] and that of the LBJ estimator was calculated by the resampling method with 500 resamplings with the perturbation variables generated from the standard exponential distribution. We keep the bandwidth $b = n^{-1/5}$ for the WLS method as in Yu et al. [26] and bandwidth $b = 4sd\left(\hat{\beta}^{0T} \mathbf{X}_i\right) n^{-1/3}$ for the LBJ method, where $sd\left(\hat{\beta}^{(0)T} \mathbf{X}_i\right)$ is the standard deviation of $\hat{\beta}^{(0)T} \mathbf{X}_i$ based on the initial estimator $\hat{\beta}^{(0)}$.

For the proposed test, we use the setting 3 to investigate the type I error and power of the test. For the power of the test, we use the following hypotheses:

$H_0 : \mathbf{\Lambda}\beta = a \quad \text{versus} \quad H_1 : \mathbf{\Lambda}\beta \neq a$, where
$\mathbf{\Lambda} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad a = \begin{bmatrix} -0.97 \\ 0.97 \end{bmatrix}$, which is equivalent to test if $\beta_1 = -0.97$ and $\beta_3 = 0.97$. For the type I error of the test, we use the following hypotheses:
$H_0 : [0 \quad 1 \quad 1 \quad 0 \quad 1]\,\beta = 0 \text{ versus } H_1 : [0 \quad 1 \quad 1 \quad 0 \quad 1]\,\beta \neq 0$, which is equivalent to test if $\beta_1 + \beta_2 + \beta_4 = 0$.

5.3. Simulation results

Tables 1–4 summarize the results from the WLS method and the LBJ method based on 500 Monte Carlo runs. The scenarios 1, 2, 4 with σ_1^2, which satisfies the assumption in model (1), are same as those in Pang et al. [28] and therefore we copied the results directly from them for these scenarios. In the tables, *bias* is the bias averaged from the 500 Monte Carlo runs, *sd* is the empirical standard deviation, *se* is the estimated standard deviation and *cov* is the coverage percentage of 95% confidence interval.

Table 1. Scenario 1: Homoscedastic error and covariate independent censoring.

			WLS				LBJ			
n	CP	Coefficient	Bias	sd	se	cov	Bias	sd	se	cov
					Normal error					
200	20%	β_1	−0.006	0.083	0.107	0.940	0.004	0.095	0.096	0.958
		β_2	−0.001	0.091	0.116	0.948	0.002	0.110	0.112	0.934
	40%	β_1	−0.013	0.093	0.121	0.942	0.002	0.102	0.107	0.948
		β_2	−0.007	0.103	0.129	0.936	0.003	0.122	0.124	0.944
400	20%	β_1	−0.002	0.056	0.072	0.948	0.005	0.068	0.070	0.932
		β_2	−0.004	0.060	0.078	0.962	0.006	0.078	0.082	0.942
	40%	β_1	−0.019	0.066	0.081	0.934	0.007	0.079	0.079	0.948
		β_2	−0.016	0.071	0.087	0.946	0.005	0.094	0.091	0.948
					Extreme error					
200	20%	β_1	0.001	0.104	0.138	0.932	0.006	0.094	0.094	0.954
		β_2	−0.013	0.120	0.148	0.930	0.003	0.112	0.109	0.942
	40%	β_1	−0.008	0.120	0.154	0.928	0.006	0.107	0.106	0.948
		β_2	−0.024	0.138	0.167	0.940	0.003	0.130	0.123	0.944
400	20%	β_1	−0.014	0.073	0.097	0.944	0.003	0.066	0.066	0.952
		β_2	−0.006	0.081	0.103	0.944	0.002	0.076	0.077	0.954
	40%	β_1	−0.026	0.090	0.108	0.926	0.004	0.077	0.075	0.930
		β_2	−0.019	0.095	0.115	0.940	0.007	0.088	0.087	0.926

Table 2. Scenario 2: Heteroscedastic error and covariate independent censoring.

n	CP	Coefficient	WLS				LBJ			
			Bias	sd	se	cov	Bias	sd	se	cov
Normal error										
200	20%	β_1	0.003	0.046	0.063	0.958	0.015	0.117	0.124	0.928
		β_2	0.006	0.054	0.071	0.948	0.009	0.116	0.122	0.936
	40%	β_1	0.007	0.053	0.077	0.978	0.034	0.126	0.134	0.924
		β_2	0.011	0.064	0.084	0.952	0.025	0.130	0.133	0.920
400	20%	β_1	0.002	0.029	0.04	0.96	0.015	0.084	0.091	0.946
		β_2	0.002	0.037	0.047	0.942	0.012	0.082	0.089	0.954
	40%	β_1	0.005	0.035	0.048	0.964	0.026	0.090	0.099	0.940
		β_2	0.008	0.045	0.055	0.944	0.019	0.087	0.097	0.946
Extreme error										
200	20%	β_1	−0.007	0.059	0.084	0.944	0.012	0.121	0.124	0.936
		β_2	−0.012	0.072	0.094	0.938	0.003	0.119	0.121	0.924
	40%	β_1	0.001	0.068	0.104	0.956	0.024	0.129	0.131	0.946
		β_2	−0.006	0.085	0.111	0.932	0.012	0.129	0.130	0.948
400	20%	β_1	−0.011	0.038	0.055	0.950	0.001	0.085	0.088	0.944
		β_2	−0.008	0.047	0.062	0.942	0.001	0.082	0.085	0.944
	40%	β_1	−0.005	0.045	0.067	0.948	0.008	0.092	0.094	0.940
		β_2	−0.002	0.055	0.073	0.958	0.002	0.089	0.091	0.950

Table 3. Scenario 3: Heteroscedastic error and covariate-independent censoring with 4 covariates.

n	CP	Variance		WLS Bias	WLS sd	WLS se	WLS cov	LBJ Bias	LBJ sd	LBJ se	LBJ cov
200	20%	σ_1^2	β_1	−0.003	0.012	0.016	0.954	−0.004	0.051	0.061	0.934
			β_2	0.004	0.016	0.022	0.962	−0.003	0.066	0.077	0.92
			β_3	0.001	0.014	0.017	0.940	−0.003	0.046	0.055	0.944
			β_4	−0.001	0.013	0.017	0.962	0.006	0.049	0.054	0.922
			Type I		0.048				0.048		
			Power		0.608				0.102		
		σ_2^2	β_1	−0.001	0.073	0.097	0.966	−0.001	0.009	0.011	0.940
			β_2	0.000	0.082	0.11	0.960	0.000	0.012	0.014	0.924
			β_3	0.000	0.072	0.142	0.970	−0.001	0.008	0.010	0.964
			β_4	0.001	0.071	0.137	0.946	0.001	0.009	0.010	0.932
			Type I		0.054				0.072		
			Power		0.878				0.886		
		σ_3^2	β_1	−0.001	0.086	0.122	0.960	−0.001	0.008	0.009	0.904
			β_2	0.002	0.087	0.125	0.956	0.000	0.007	0.009	0.924
			β_3	0.012	0.097	0.236	0.954	−0.001	0.008	0.01	0.934
			β_4	0.012	0.1	0.234	0.944	0.000	0.008	0.009	0.938
			Type I		0.048				0.078		
			Power		0.652				0.986		
	40%	σ_1^2	β_1	−0.004	0.016	0.022	0.954	−0.007	0.060	0.071	0.928
			β_2	0.006	0.021	0.031	0.960	0.000	0.073	0.083	0.928
			β_3	0.001	0.018	0.023	0.950	−0.002	0.054	0.063	0.934
			β_4	−0.001	0.016	0.023	0.964	0.007	0.057	0.063	0.916
			Type I		0.052				0.064		
			Power		0.398				0.114		
		σ_2^2	β_1	−0.001	0.082	0.111	0.964	−0.001	0.011	0.013	0.938
			β_2	0.001	0.091	0.126	0.952	0.000	0.013	0.015	0.914
			β_3	0.000	0.080	0.170	0.974	0.000	0.009	0.011	0.946
			β_4	0.001	0.082	0.164	0.950	0.002	0.011	0.012	0.928
			Type I		0.046				0.078		
			Power		0.774				0.798		
		σ_3^2	β_1	−0.002	0.092	0.135	0.956	−0.001	0.009	0.01	0.898
			β_2	0.002	0.098	0.141	0.956	0.000	0.009	0.01	0.916
			β_3	0.013	0.109	0.255	0.954	−0.001	0.01	0.011	0.924
			β_4	0.014	0.107	0.249	0.952	0.000	0.009	0.01	0.926
			Type I		0.042				0.062		
			Power		0.478				0.938		
400	20%	σ_1^2	β_1	−0.001	0.009	0.011	0.944	−0.001	0.037	0.045	0.948
			β_2	0.002	0.011	0.014	0.956	0.003	0.046	0.055	0.940
			β_3	0.001	0.009	0.011	0.948	0.002	0.030	0.040	0.968
			β_4	−0.001	0.008	0.011	0.964	−0.004	0.031	0.039	0.960
			Type I		0.054				0.050		
			Power		0.874				0.142		
		σ_2^2	β_1	0.000	0.005	0.006	0.946	0.000	0.007	0.008	0.938
			β_2	0.000	0.006	0.007	0.948	0.000	0.008	0.010	0.948
			β_3	0.000	0.004	0.007	0.938	0.000	0.005	0.007	0.962
			β_4	0.000	0.004	0.007	0.944	−0.001	0.006	0.007	0.944
			Type I		0.052				0.058		
			Power		1				0.986		
		σ_3^2	β_1	0.000	0.006	0.010	0.978	0.000	0.005	0.006	0.934
			β_2	0.001	0.006	0.010	0.956	0.000	0.005	0.006	0.946
			β_3	0.007	0.011	0.048	0.966	0.000	0.005	0.007	0.940
			β_4	0.005	0.010	0.047	0.974	−0.001	0.005	0.007	0.956
			Type I		0.034				0.062		
			Type I		0.034				0.062		
			Power		0.880				1		
	40%	σ_1^2	β_1	−0.002	0.011	0.014	0.934	−0.002	0.044	0.052	0.956
			β_2	0.004	0.014	0.02	0.958	0.005	0.049	0.059	0.946

	β_3	0.002	0.012	0.015	0.950	0.004	0.035	0.045	0.964
	β_4	−0.002	0.011	0.015	0.968	−0.006	0.036	0.045	0.952
	Type I		0.064				0.054		
	Power		0.7				0.116		
σ_2^2	β_1	−0.001	0.006	0.007	0.962	0.000	0.008	0.009	0.950
	β_2	0.001	0.007	0.009	0.964	0.000	0.009	0.011	0.944
	β_3	0.000	0.005	0.011	0.950	0.000	0.006	0.008	0.962
	β_4	0.000	0.005	0.010	0.954	0.000	0.007	0.009	0.940
	Type I		0.046				0.054		
	Power		1				0.958		
σ_3^2	β_1	−0.001	0.007	0.012	0.964	0.000	0.006	0.007	0.938
	β_2	0.001	0.007	0.013	0.962	0.001	0.006	0.007	0.944
	β_3	0.008	0.013	0.057	0.966	0.000	0.006	0.008	0.932
	β_4	0.006	0.012	0.055	0.970	−0.001	0.006	0.007	0.942
	Type I		0.032				0.072		
	Power		0.778				0.998		

Table 4. Scenario 4: Heteroscedastic error and covariate-dependent censoring.

			WLS				LBJ			
n	CP	Coefficient	Bias	sd	se	cov	Bias	sd	se	cov
					Normal error					
200	20%	β_1	−0.001	0.05	0.069	0.960	0.016	0.117	0.124	0.926
		β_2	0.008	0.053	0.073	0.956	0.000	0.123	0.129	0.938
	40%	β_1	0.001	0.057	0.085	0.958	0.036	0.124	0.128	0.926
		β_2	0.008	0.059	0.085	0.946	0.020	0.128	0.134	0.936
400	20%	β_1	0.004	0.032	0.044	0.952	0.009	0.085	0.096	0.948
		β_2	0.002	0.037	0.048	0.962	0.001	0.082	0.096	0.972
	40%	β_1	0.009	0.037	0.051	0.958	0.024	0.088	0.101	0.940
		β_2	0.004	0.041	0.054	0.944	0.013	0.088	0.105	0.956
					Extreme error					
200	20%	β_1	−0.007	0.062	0.09	0.952	0.011	0.114	0.121	0.966
		β_2	−0.012	0.071	0.096	0.938	0.003	0.113	0.120	0.962
	40%	β_1	−0.001	0.077	0.111	0.950	0.027	0.125	0.128	0.946
		β_2	−0.006	0.081	0.112	0.950	0.010	0.124	0.129	0.954
400	20%	β_1	−0.01	0.041	0.058	0.942	0.005	0.076	0.087	0.956
		β_2	−0.005	0.046	0.064	0.952	0.002	0.078	0.085	0.954
	40%	β_1	−0.006	0.049	0.07	0.944	0.013	0.083	0.094	0.964
		β_2	−0.003	0.054	0.073	0.938	0.000	0.085	0.093	0.946

As seen from Table 1, for scenario 1 where the data is homoscedastic, the WLS and the LBJ are asymptotically equivalent confirming with Theorem 4.1. For finite sample, the WLS method is more efficient than the LBJ method when data follows normal distribution, while the LBJ method is more efficient when the data follows extreme value distribution. It suggests that the nonparametric smoothing estimation works well for homoscedastic data.

As demonstrated in Table 2, for scenario 2 where the data is heteroscedastic, the WLS method is more efficient than the LBJ method, whether the data follow normal distribution or extreme value distribution and in spite of the censoring rate and sample size. Again it confirms with Theorem 4.1.

The results from scenario 3 are summarized in Table 3. When the variance function σ_1^2 is used and the variance function satisfies the assumption in model (1), the WLS is always more efficient than the LBJ as observed in scenario 2. When the variance function does not satisfy the assumption, such as σ_2^2 or σ_3^2, the WLS method loses more efficiency than the LBJ method, especially for small sample sizes and the domain of the variance function is much different from μ, such as σ_3^2. We note that the LBJ is more efficient than the WLS method for $n = 200$ with either variance function σ_2^2 or σ_3^2. When $n = 400$, the efficiency of these two methods is similar for σ_2^2, but the LBJ is still more efficient than the WLS method for σ_3^2. This indicates the LBJ is less affected by the violation of the assumption of the variance function. For the aspect of estimated variance, it matches the empirical variance well for the LBJ method, however, it always over estimates the empirical variance for the WLS method. This may be due to the small number of bootstrap samples used, compared with sample size 200 or 400. When the sample size is 100 in the original paper [26], the estimated variance is smaller than the empirical variance for some cases. For the coverage probability of the confidence interval, both methods perform reasonably well, and therefore the results from both methods are valid. This provides a guideline for real data analysis, that is, we can apply both methods and use the more efficient one to do inference.

For the type I error of the new proposed tests, the WLS-based test is a little conservative than the LBJ-based test, but both of them have type I error close to 0.05. For the power of the tests, they are associated with the estimation efficiency. Specifically, when σ_1^2 is used, the WLS estimation is more efficient than the LBJ estimation. Then the WLS-based test is more powerful than the LBJ-based test. When σ_2^2 is used and sample size is 200, the WLS estimation is less efficient than the LBJ estimation. We observed that the WLS-based test is less powerful than the LBJ-based test. For the cases with σ_2^2 and sample size 400, the WLS-based test is more powerful than the LBJ-based test as the WLS estimation is more efficient than the LBJ estimation. When σ_3^2 is used, the WLS estimation is less efficient than the LBJ estimation. Then the WLS-based test is less powerful than the LBJ-based test.

The results for scenario 4 in Table 4 agree with those from the scenario 2, which indicates the WLS is more efficient than the LBJ as well when the censoring depends on the covariates and the assumption of the variance function is satisfied.

As both methods use the Kaplan–Meier estimation, which results in discrete estimating equations. Therefore, the convergence of the algorithm is not guaranteed and the iterative sequence may become trapped in a loop, oscillating between two or more points [14]. We investigated the non-convergence rate of these two methods. We use scenario 3 with $n = 200$ and 20% censoring. Both failed to converge only once in 500 Monte Carlo runs.

6. Real data analyses

6.1. Stanford heart transplant data

We first consider the Stanford heart transplant data [32], which has 184 observations. Two covariates are provided in this dataset, T5 mismatch score and age. We use the model in Miller and Halpern [32].

Specifically, we delete the observations with missing values in T5 mismatch score, which leads to a dataset with 157 observations. Among them, 102 are uncensored. Then we fit Miller and Halpern's [32] model by regressing the base-10 logarithm of the survival time on the patient's age (in years) and age^2 as follows: $\log_{10} T = \beta_0 + \beta_1 age + \beta_2 age^2 + \epsilon$. Note that this is the best model that Miller and Halpern got after they figured out that T5 mismatch score is not significant and age^2 is significant. The age ranges from 16 years to 64 years, with mean 41 years. The variance model we used is $E(\epsilon^2) = \sigma^2(\mu) + \phi$, where ϕ is random error, $\sigma^2(.)$ is an unspecified function, and $\mu = \beta_0 + \beta_1 age + \beta_2 age^2$. Note that we did not specify any particular form for the variance function of ϵ.

The WLS method, LBJ method and Buckley–James (BJ) method were applied for this data. The bootstrapping size in the WLS method and the resampling size in the LBJ method are the same as in the simulation studies, which are 50 and 500, respectively. The results are summarized in Table 5. We can see from the table that the variances of the parameter estimators are much larger for the WLS method than for the LBJ method. This may be because the assumption of the variance function is not satisfied and the sample size is small, according to the investigation through simulations. We checked the heteroscedasticity of the data using the estimated variance function by the WLS method as shown in Figure 1. It supports that the data is heteroscedastic. As both WLS and LBJ methods are valid, we should use the estimation from the LBJ method for this particular data.

6.2. Primary biliary cirrhosis (PBC) data

Second, we consider Mayo Clinic Primary Biliary Cirrhosis (PBC) Data [33]. This data is from the clinic trial conducted by Mayo Clinic in PBC between 1974 and 1984. The goal of this data is to investigate the effects of factors on the survival of PBC patients. There are totally 418 observations. We consider the covariates of age (in years), hepato ($1 =$presence of hepatomegaly or enlarged liver, $0 =$ no such presence), stage (histologic stage of disease, takes values 1 to 4), edema ($0 =$no edema, $0.5 =$untreated or successfully treated, $1 =$ edema despite diuretic therapy). After deleting observations with missing values in variable hepato, we have 312 observations and among them, 125 are uncensored. As in Wei et al. [1], we take base 10 logarithm of the survival time, then the model is constructed as follows: $\log_{10} T = \beta_0 + \beta_1 age + \beta_2 hepato + \beta_3 stage + \beta_4 edema + \epsilon,$ (3) and the variance model is $E(\epsilon^2) = \sigma^2(\mu) + \phi$, where ϕ is random error, $\sigma^2(.)$ is an unspecified function, and $\mu = \beta_0 + \beta_1 age + \beta_2 hepato + \beta_3 stage + \beta_4 edema$.

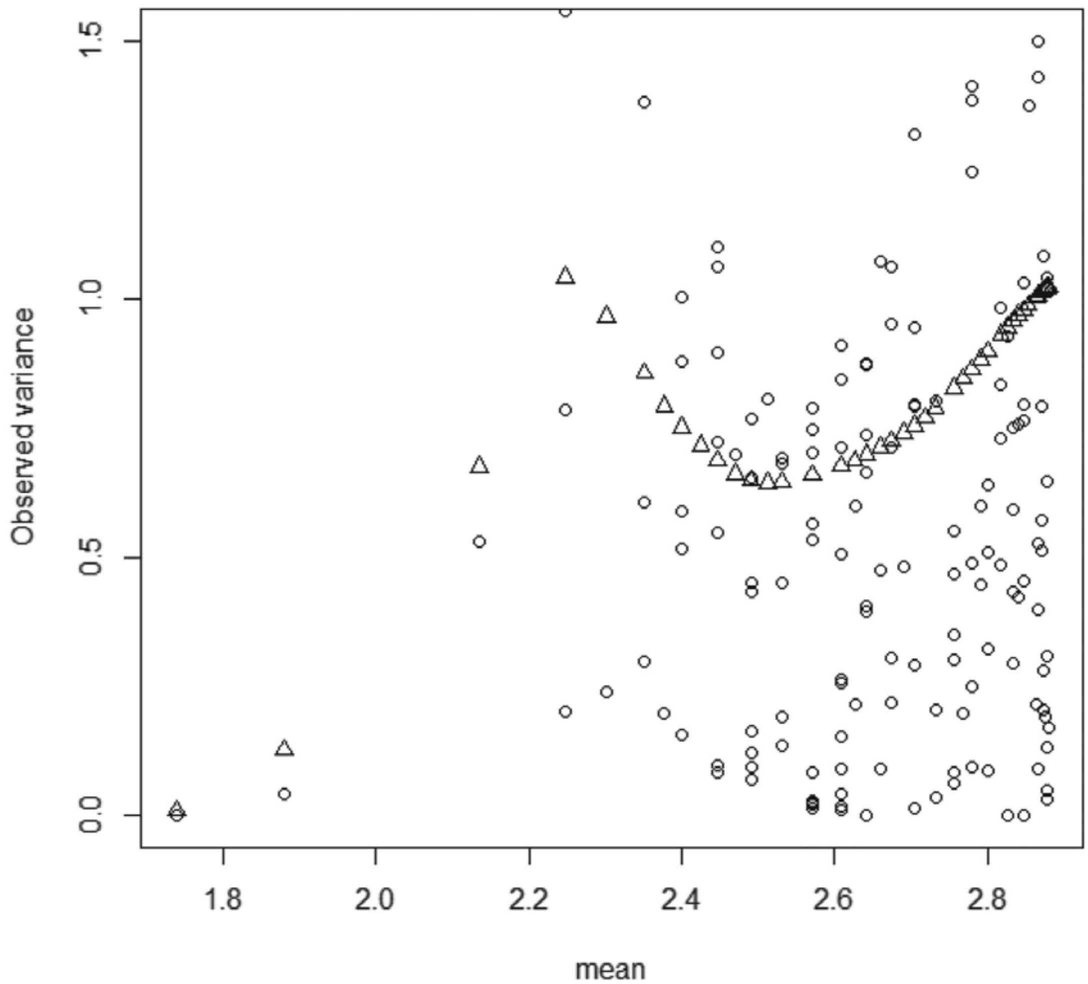

Figure 1. Variance function estimated from the weighted least squares method for Stanford heart transplant dataset. The dots are the observed variances and the triangles are the estimated variances.

Table 5. Analysis of the PBC dataset.

	BJ			LBJ			WLS		
	Estimate	SD	95% CI	Estimate	SD	95% CI	Estimate	SD	95% CI
Age	−0.009	0.003	(−0.014,−0.003)	−0.007	0.002	(−0.012,−0.002)	−0.005	0.002	(−0.010,−0.001)
Hepato	−0.194	0.067	(−0.325,−0.062)	−0.149	0.055	(−0.257,−0.041)	−0.131	0.050	(−0.230,−0.032)
Stage	−0.188	0.046	(−0.276,−0.098)	−0.152	0.036	(−0.222,−0.081)	−0.115	0.030	(−0.173,−0.057)
Edema	−0.789	0.113	(−1.011,−0.566)	−0.768	0.125	(−1.013,−0.523)	−0.843	0.133	(−1.103,−0.583)

We applied the BJ method [9], WLS method [26] and LBJ method [28] to this data. The results are summarized in Table 6. We can see that the WLS method is most efficient except for the covariate edema. This indicates that the data is heteroscedastic and the variance function may depend on a domain that is close to the mean values of the data in the model (3). We check the variance function based on the WLS method, which is shown in Figure 2. It clearly shows that the data is heteroscedastic. Based on the work of Yu and Peace [25], the results from the BJ method are not reliable, i.e. the estimated variances may underestimate the true variance and hence the coverage probability of the confidence interval may be much lower than the nominal level. Therefore, we can use the estimation for the first three covariates based on the WLS method and use the estimation for edema based on the LBJ method because both of them have reasonable coverage probability of the confidence interval.

In model (3), we may be interested in whether hepoto has the same effect as stage on the survival. Then we can apply the new proposed hypotheses test, i.e.
$H_0 : \beta_2 = \beta_3 \quad \textbf{versus} \quad H_1 : \beta_2 \neq \beta_3$. The test based on both the LBJ and WLS methods do not reject H_0, and hence we conclude that hepoto and stage have similar effect on the survival time of PBC patients.

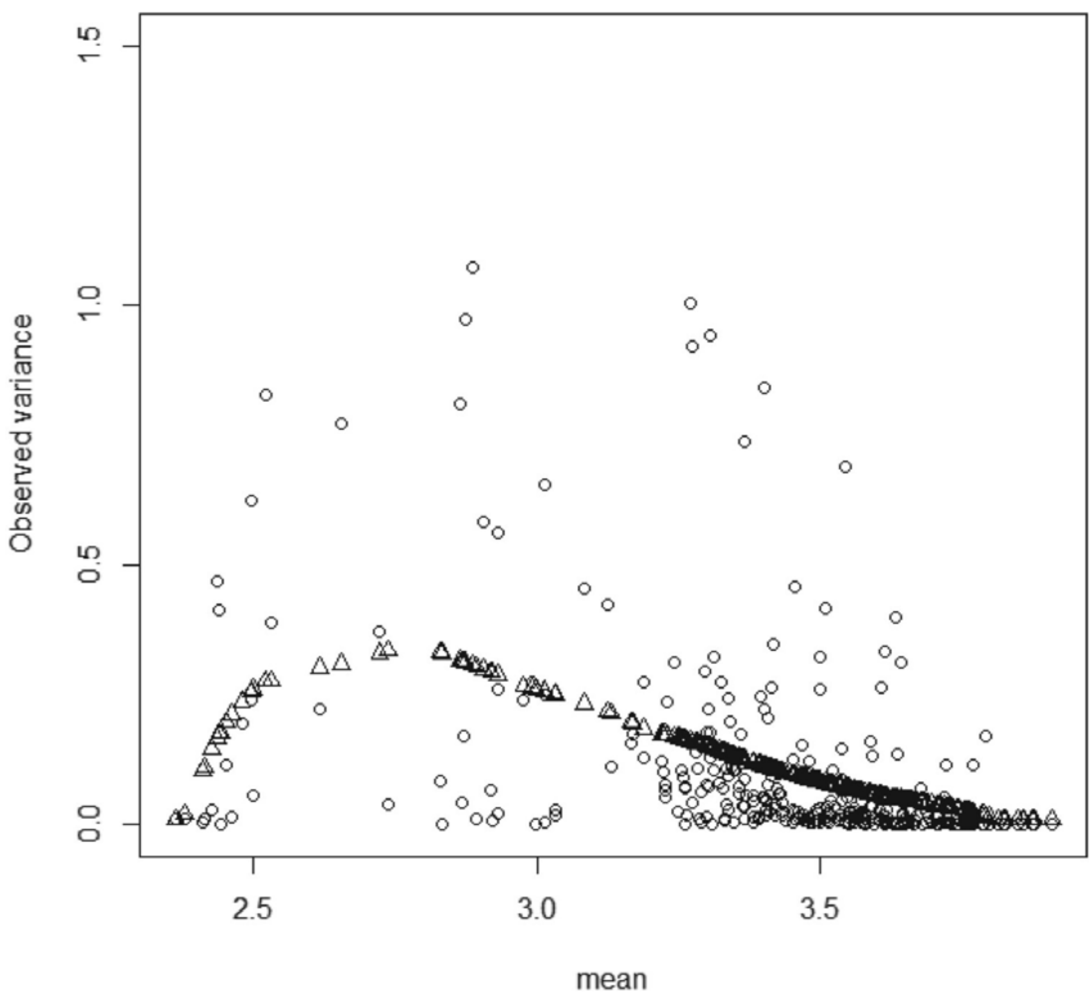

Figure 2. Variance function estimated from the weighted least squares method for PBC dataset. The dots are the observed variances and the triangles are the estimated variances.

Table 6. Analysis of the PBC dataset.

	BJ			LBJ			WLS		
	Estimate	SD	95% CI	Estimate	SD	95% CI	Estimate	SD	95% CI
age	0.111	0.048	(0.016, 0.206)	0.110	0.043	(0.025, 0.195)	0.076	0.083	(−0.087, 0.239)
age^2	−0.002	0.001	(−0.003, −0.001)	−0.002	0.001	(−0.003, −0.001)	−0.001	0.001	(−0.003, 0.001)

7. Discussion and conclusion

This paper compared the WLS method and the LBJ method. Both theoretical and numerical investigations showed that the WLS method is more efficient when data is heteroscedastic, while they are similar when data is homoscedastic, under the assumption that the variance is a function of the mean of the data. In addition, we investigate their performance when the assumption is violated, i.e. the variance function is not a function of the mean, instead, it is a function of a subset of the covariates. It shows the LBJ method is more efficient for some scenarios such as the domain of the variance function is far away from the mean of the data or when the sample size is small. Then we provide a guideline for real data analysis. We suggest to apply both methods for real data analysis and select the one with the more efficient inference, because we observed through simulations that both methods are valid inference methods in terms of the coverage probability of the confidence interval. Then we can obtain the most efficient inference for real data analysis based on current methodologies.

Although they [26,28] are the main methods for heteroscedastic survival data so far, both methods are based on Kaplan–Meier estimate, which is discrete and time-consuming, especially for data with heavy censoring and large sample sizes. Therefore, they may not be able to be applied to such situations. Moreover, both methods did not consider survival data with high-dimensional covariates. Further research can be conducted to overcome these drawbacks.

In addition, we developed a new test for testing effects of any subsets of the parameters, which makes the inference procedure for heteroscedastic survival data complete. For future research, we can extend both the WLS method and the LBJ method to more complex settings, such as heteroscedastic data with frailty, with measurement error, or with cure rate models, etc., and we are actively investigating these extensions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are openly available in Miller and Halpern [32] and Fleming and Harrington [33].

References

1. Wei LJ, Ying Z, Lin DY. Linear-regression analysis of censored survival-data based on rank-tests. Biometrika. 1990;77:845–851. doi: 10.1093/biomet/77.4.845

2. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: Wiley; 2002.

3. Cox DR. Regression models and life tables (with discussion). J R Stat Series B. 1972;34:187–220.

4. Tsiatis AA. Estimating regression parameters using linear rank-tests for censored-data. Ann Stat. 1990;18:354–372. doi:10.1214/aos/1176347504

5. Lai TL, Ying ZL. Rank regression methods for left-truncated and right-censored data. Ann Stat. 1991;19:531–556.

6. Robins J, Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. Biometrika. 1992;79:311–319.

7. Ying ZL. A large-sample study of rank estimation for censored regression data. Ann Stat. 1993;21:76–99.

8. Lin DY, Ying ZL. Semiparametric inference for the accelerated life model with time-dependent covariates. J Stat Plan Inference. 2000;44:47–63. doi:10.1016/0378-3758(94)00039-X

9. Jin ZZ, Lin DY, Wei LJ, et al. Rank-based inference for the accelerated failure time model. Biometrika. 2003;90:341–353. doi:10.1093/biomet/90.2.341

10. Zhou M. Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. Biometrika. 2008;92:492–498. doi:10.1093/biomet/92.2.492

11. Buckley J, James I. Linear-regression with censored data. Biometrika. 1979;66:429–436. doi:10.1093/biomet/66.3.429

12. Ritov Y. Estimation in a linear-regression model with censored-data. Ann Stat. 1990;18:303–328.

13. Lai TL, Ying ZL. Large sample theory of a modified Buckley–James estimator for regression-analysis with censored-data. Ann Stat. 1991;19:1370–1402.

14. Jin ZZ, Lin DY, Ying ZL. On least-squares regression with censored data. Biometrika. 2006;90:147–161. doi:10.1093/biomet/93.1.147

15. Zeng D, Lin DY. Efficient estimation for the accelerated failure time model. J Am Stat Assoc. 2007;102:1387–1396. doi:10.1198/016214507000001085

16. Stare J, Heinzl H, Harrell F. On the use of Buckley and James least squares regression for survival data. New Approaches Appl Stat. 2000;16:125–134.

17. Zhang M, Davidian M. Smooth semiparametric regression analysis for arbitrarily censored time-to-event data. Biometrics. 2008;64:567–576. doi:10.1111/biom.2008.64.issue-2

18. Chen S, Khan S. Estimating censored regression models in the presence of nonparametric multiplicative heteroscedasticity. J Econom. 2000;98:283–316. doi:10.1016/S0304-4076(00)00020-8

19. Zhou M, Bathke A, Kim M. Empirical likelihood analysis for the heteroscedastic accelerated failure time model. Stat Sin. 2012;22:295–316. doi:10.5705/ss.2011.v22n1a

20. Heuchenne C, Van Keilegom I. Polynomial regression with censored data based on preliminary nonparametric estimation. Ann Inst Stat Math. 2007;59:273–297. doi:10.1007/s10463-006-0066-4

21. Liu W, Lu X. Weighted least squares method for censored linear models. J Nonparametr Stat. 2009;21:787–799. doi:10.1080/10485250902795636

22. Fan J, Gijbels I. Censored regression: local linear approximations and their applications. J Am Stat Assoc. 1994;89:560–570. doi:10.1080/01621459.1994.10476781

23. Yu L. Nonparametric quasi-likelihood for right censored data. Lifetime Data Anal. 1993;17:594–607. doi:10.1007/s10985-011-9200-6

24. Yu L, Yu R, Liu L, et al. Extended quasi-likelihood in the generalized linear model for right-censored data. Stat Med. 2012;31:1369–1379. doi:10.1002/sim.v31.13

25. Yu L, Peace EK. Spline nonparametric quasi-likelihood regression within the frame of the accelerated failure time model. Comput Stat Data Anal. 2012;56:2675–2687. doi:10.1016/j.csda.2012.02.009

26. Yu L, Liu L, Chen D. Weighted least-squares method for right-censored data in accelerated failure time model. Biometrics. 2013;69:358–365. doi:10.1111/biom.v69.2

27. Yu L, Liu L, Chen D. A homoscedasticity test for the accelerated failure time model. Comput Stat. 2019;34:433–446. doi:10.1007/s00180-018-0840-9

28. Pang L, Lu W, Wang H. Local Buckley-James estimation for heteroscedastic accelerated failure time model. Stat Sin. 2015;25:863–877.

29. Choi B, Fine J, Brookhart A. On two-stage estimation of structural instrumental variable models. Biometrika. 2017;104:881–899. doi:10.1093/biomet/asx056

30. Fan J, Gijbels I. Local polynomial modelling and its applications. London: Chapman and Hall/CRC; 1996.

31. Nadaraya E. On estimating regression. Theory Probab Its Appl. 1964;9:141–142. doi:10.1137/1109020

32. Miller RG, Halpern J. Regression with censored data. Biometrika. 1982;69:521–531. doi:10.1093/biomet/69.3.521

33. Fleming TR, Harrington DP. Counting processes and survival analysis. New York: Wiley; 1991.

34. Lo SH, Singh K. The product-limit estimator and the bootstrap: some asymptotic representations. Probab Theory Relat Fields. 1986;71:455–465. doi:10.1007/BF01000216

Appendix

Proof of Theorem 4.1

Define $\overline{\mathbf{U}}^*\left(\beta,\sigma^2\left(\mu_i\right)\right)=\sum_{i=1}^n\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\widetilde{E}\left(e\middle|e>r_i\right)/\sigma\left(\mu_i\right)$, where $\overline{\mathbf{X}}=n^{-1}\sum_{i=1}^n\mathbf{X}_i$, $r_i=\left(y_i-\beta^T\mathbf{X}_i\right)/\sigma\left(\mu_i\right)$ and suppose $\bar{\beta}^*$ is the solution of $\overline{\mathbf{U}}^*\left(\beta,\sigma^2\left(\mu_i\right)\right)=0$;
$\overline{\mathbf{U}}\left(\beta,\sigma^2\left(\mu_i\right)\right)=\sum_{i=1}^n\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\sigma\left(\mu_i\right)\widetilde{E}\left(e\middle|e>r_i\right)$ and $\bar{\beta}$ is the solution of $\overline{\mathbf{U}}\left(\beta,\sigma^2\left(\mu_i\right)\right)=0$.
Further,
$V_n\left(\beta\right)=\sum_{i=1}^n\left\{\int_{-\infty}^{\infty}t\,\mathrm{d}E\left(Y_i^x\left(t\right)\right)+\int_{-\infty}^{\infty}\int_t^{\infty}\frac{1-F_i^\epsilon\left(s\right)}{1-F_i^\epsilon\left(t\right)}\,\mathrm{d}s\,\mathrm{d}E\left(J_i^x\left(t\right)\right)\right\}$, where F^ϵ is the true density function for ϵ, $Y_i^x\left(t\right)=\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)I\{\hat{\epsilon}_i(\beta)\geq t\}$ and $J_i^x\left(t\right)=\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)I\{\hat{\epsilon}_i(\beta)\geq t,\delta_i=0\}$, where $\hat{\epsilon}_i\left(\beta\right)=y_i-\beta^T\mathbf{X}_i$.

According to Gauss–Markov theorem, when the data is heteroscedastic, $\bar{\beta}^*$ is the efficient estimator based on the data $\delta_i y_i+\left(1-\delta_i\right)\left(\beta^T\mathbf{X}_i+\sigma_i\left(\mu_i\right)\widetilde{E}\left(e_i\middle|e_i>r_i\right)\right),i=1,\ldots,n$. Clearly, $var\left(\bar{\beta}\right)-var\left(\bar{\beta}^*\right)$ is positive definite. When the data is homoscedastic, $var\left(\bar{\beta}\right)$ and $var\left(\bar{\beta}^*\right)$ have the same efficiency.

Yu et al. [26] showed that $\bar{\beta}^*$ and $\hat{\beta}$ have the same limiting distribution. Therefore, it is sufficient to prove that $var\left(\hat{\beta}\right)-var\left(\bar{\beta}\right)$ is nonnegative definite. Based on the work of Pang et al. [28], $var\left(\hat{\beta}\right)=n^{-1}\mathbf{D}^{-1}\mathbf{H}\mathbf{D}^{-1}$ and based on the work of Lai and Ying [13], $var\left(\bar{\beta}\right)=n^{-1}\mathbf{D}^{-1}\overline{\mathbf{H}}\mathbf{D}^{-1}$, where $\mathbf{D}=\mathrm{d}V_n(\beta)/\mathrm{d}\beta$, $\overline{\mathbf{H}}=var(\overline{\mathbf{U}}(\beta,\sigma^2(\boldsymbol{\mu})))$ and $\mathbf{H}=var(\mathbf{L}(\beta))$. Because $var\left(\hat{\beta}\right)-var\left(\bar{\beta}\right)=n^{-1}\mathbf{D}^{-1}\left(\mathbf{H}-\overline{\mathbf{H}}\right)\mathbf{D}^{-1}$, it is sufficient if we can show $\overline{\mathbf{U}}(\beta,\sigma^2(\boldsymbol{\mu}))$ is asymptotically equivalent to $\mathbf{L}(\beta)$. Now we will show that $\overline{\mathbf{U}}(\beta,\sigma^2(\boldsymbol{\mu}))$ can be expressed as the sum of three parts, which are equivalent in probability to those of $\mathbf{L}(\beta)$ in Pang et al. [28]. Specifically,
$\overline{\mathbf{U}}(\beta,\sigma^2(\boldsymbol{\mu}))=u_1(\beta,\sigma^2(\boldsymbol{\mu}))+u_2(\beta,\sigma^2(\boldsymbol{\mu}))+u_3(\beta,\sigma^2(\boldsymbol{\mu}))+o_p(1)$, where

$$u_1=-\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\left\{\int_{-\infty}^{\infty}t\,\mathrm{d}Y_i\left(t\right)+\int_{-\infty}^{\infty}\int_t^{\infty}\frac{1-F_i^\epsilon\left(s\right)}{1-F_i^\epsilon\left(t\right)}\,\mathrm{d}s\,\mathrm{d}J_i\left(t\right)\right\},$$

$$u_2=-\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\int_{-\infty}^{\infty}\int_t^{\infty}\frac{1-F_i^\epsilon\left(s\right)}{\{1-F_i^\epsilon(t)\}^2}\left\{\widehat{F}_i^\epsilon\left(t\right)-F_i^\epsilon\left(t\right)\right\}\mathrm{d}s\,\mathrm{d}J_i\left(t\right),$$ By

$$u_3=-\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\int_{-\infty}^{\infty}\int_t^{\infty}\frac{\widehat{F}_i^\epsilon\left(s\right)-F_i^\epsilon\left(s\right)}{1-F_i^\epsilon\left(t\right)}\,\mathrm{d}s\,\mathrm{d}J_i\left(t\right).$$

Theorem 1 of Lo and Singh [34],

$$\widehat{F}_i\left(t\right)-F_i\left(t\right)=\frac{1}{n}\sum_{j=1}^n\xi\left(y_j,\delta_j,t,\mu_i\right)+o_p\left(n^{-1/2}\right),$$ and hence

$$\widehat{F}_i^\epsilon\left(t\sigma\left(\mu_i\right)\right)-F_i^\epsilon\left(t\sigma\left(\mu_i\right)\right)=\frac{1}{n}\sum_{j=1}^n f_\mu\left(\mu_i\right)\xi\left(y_j,\delta_j,t\sigma\left(\mu_i\right),\mu_i\right)+o_p\left(n^{-1/2}\right),$$ where

$f^\epsilon(.)$ and $f_\mu(.)$ are densities for ϵ and μ respectively. Then it is obvious $1/\sqrt{n}\left(\overline{\mathbf{U}}(\beta,\sigma^2(\boldsymbol{\mu}))-\mathbf{L}(\beta)\right)=o_p(1)$. Then the proof is complete.