Simulation Study of Profile Likelihood Method in the Accelerated Failure Time Model

Yinding Wang* Jiajia Zhang†

Abstract
The proportional hazards (PH) model and the accelerated failure time (AFT) model are the most popular models to fit the censoring data. Sometimes, the PH assumption cannot be satisfied, and therefore, the AFT model becomes a useful alternative to the PH model. Comparing with the least square method and the rank estimation method, the profile likelihood method needs less computation time in estimating the parameters and associated variances in the semi-parametric AFT model. The main aim of this paper is to investigate the impact of the different bandwidths on the estimations in the profile likelihood method. Additionally, we also evaluate its performances according to the different sample sizes and censoring rates. Finally, we make recommendation of its usage in practice. For illustration, we apply the semi-parametric AFT model to the real dataset: methadone maintenance treatment, using the profile likelihood method.

Key Words: Accelerated failure time model, Proportional hazards model, Profile likelihood, Censored data, Semi-parametric, Survival analysis

1. Introduction

The proportional hazards (PH) model and accelerated failure time (AFT) model are the most popular models in survival analysis. The PH model assumes the regression structure on the logarithm of hazard function, while the AFT model assumes the regression structure on the time scale. When the PH assumption is not satisfied, the AFT model becomes an alternative tool.

Let $T_i = \min(T_i^*, C_i)$ denote the observed time with $i = 1, 2, ..., n$, where T_i^* denotes the failure time and C_i denotes censoring time for subject i separately. Generally, the censoring is assumed to be independent and non-informative. Let δ_i be the censoring indicator with 1 if $T_i^* \leq C_i$ and 0 if $T_i^* > C_i$. Thus, the AFT model can be written as

$$\log(T_i) = \boldsymbol{\beta}'\mathbf{x}_i^* + \epsilon_i \tag{1}$$

where $\boldsymbol{\beta}$ is a p-dimensional unknown parameters, $\mathbf{x}_i^* = (x_{i1}, x_{i2}, ..., x_{ip})'$ denotes the $p \times 1$ possible covariates for subject i, and ϵ_i is a random error independent of $\mathbf{x}_i^*$. Without the distribution assumption of ϵ_i, model (1) is called as the semi-parametric AFT model.

There are many discussions in literature about the estimation methods for the semi-parametric AFT model. The least square method was first proposed by Buckley and James (1979). Stute and Wang (1993) further investigated the least square method by using the KM weights. Based on the least square method, Huang, Ma and Xie (2006) utilized two regularization approaches: the least absolute shrinkage and selection operator method, and the threshold-gradient-directed regularization method to estimate the parameters in the AFT model. A bootstrap approach was used to estimate its variance. Jin, Lin and Ying (2006) improved the least square method for the AFT model by utilizing the Gehan rank estimator as its initial values. The resampling technique is employed to estimate the variance. Tsiatis

*Yinding Wang, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29205

†Jiajia Zhang, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29205

(1990) proposed the rank estimation method in the AFT model. The equivalence of the rank estimation method and the least square method was given by Ritov (1990). Jin, Lin, Wei and Ying (2003) simplified the estimation procedure of the rank estimation method based on the Gehan-type estimator, and extended it to other weight functions. Additionally, they introduced the resampling technique to estimate the variance of estimators. Brown and Wang (2007) proposed the induced smoothing to the Gehan-Wilcoxon weighted rank regression, which could obtain the variance directly.

Zeng and Lin (2007) proposed a profile likelihood method by approximating the profile likelihood function with a kernel function. The advantages of Zeng and Lin (2007)'s method are that both the variance and the smoothed estimator of the survival function can be easily obtained. Additionally, they proved the asymptotically normality of the estimated parameters. Their simulation results showed that the profile likelihood method was much faster than those using the rank estimation method (Jin, Lin, Wei and Ying (2003)) or the least square method (Jin, Lin and Ying (2006)). However, Zeng and Lin (2007) did not investigate the impact of different bandwidths on the survival function. Therefore, the purpose of this article is to investigate the performance of the profile likelihood method with regard to the estimators of both parameters and survival functions.

The organization of this article is as follows: Section 2 provides details of the profile likelihood method proposed by Zeng and Lin (2007). Section 3 presents a comprehensive simulation study and evaluates the performances of profile likelihood method under different situations. The profile likelihood method is applied to the real dataset in Section 4. Conclusions are summarized in Section 5.

2. Profile Likelihood Method

Let $h(\cdot)$ denote a baseline hazard function of e^{ε} and $H(\cdot)$ denote its corresponding cumulative hazard function, that is $H(t) = \int_0^t h(s)ds$. The hazard function of T in the AFT model can be written as

$$h_T(t) = h(te^{-\boldsymbol{\beta}'\mathbf{x}^*})e^{-\boldsymbol{\beta}'\mathbf{x}^*} \tag{2}$$

Accordingly, we can obtain its cumulative hazard function and survival function, which can be written as $H_T(t) = H(te^{-\boldsymbol{\beta}'\mathbf{x}^*})$ and $S_T(t) = \exp(-H_T(t))$ separately.

Given the observed dataset (t_i, δ_i, x_i^*), the observed likelihood function is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} h_T(t_i)^{\delta_i} S_T(t_i) = \prod_{i=1}^{n} [h(t_i e^{-\boldsymbol{\beta}'\mathbf{x}_i^*})e^{-\boldsymbol{\beta}'\mathbf{x}_i^*}]^{\delta_i}(e^{-H(t_i e^{-\boldsymbol{\beta}'\mathbf{x}_i^*})}) \tag{3}$$

Let $r_i(\boldsymbol{\beta}) = \log(T_i) - \boldsymbol{\beta}'\mathbf{x}_i^*$, we can express the log-likelihood function as

$$\frac{1}{n}\sum_{i=1}^{n}[-\delta_i\boldsymbol{\beta}\mathbf{x}_i^* + \delta_i \log h(e^{r_i(\boldsymbol{\beta})}) - H(e^{r_i(\boldsymbol{\beta})})] \tag{4}$$

If the distribution of e^{ε_i} is known, the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ can be obtained by maximizing the equation (4) through the Newton Raphson method. The common assumptions of e^{ε_i} are the Weibull, log-normal, and log-logistic distribution. As mentioned by Zeng and Lin (2007), without specifying the distribution of e^{ε_i}, the maximum of the equation (4) with respect to $\boldsymbol{\beta}$ and $H(\cdot)$ would lead to the nonparametric maximum likelihood estimators. However, due to the non-smoothness of $H(\cdot)$, the non-parametric maximum likelihood estimator may fail in practice.

In the remaining of this section, we will introduce the profile likelihood method proposed by Zeng and Lin (2007). First, intervals containing all $e^{r_i(\boldsymbol{\beta})}$ are partitioned into J_n equally spaced intervals, $0 \equiv t_0 < t_1 < ... < t_{J_n} \equiv M$, where M denotes an upper bound for $e^{r_i(\boldsymbol{\beta})}$ over all possible $\boldsymbol{\beta}'s$ in a bounded set. A piecewise assumption of the hazard function, denoted by $h(t)$, can be written as

$$h(t) = \sum_{k=1}^{J_n} C_k I(t \in [t_{k-1}, t_k))$$

where $I(\cdot)$ is an indicator function. The corresponding cumulative hazard function, denoted by $H(t)$, can be written as

$$H(t) = \sum_{k=1}^{J_n} C_k(t - t_k) I(t_{k-1} \leq t < t_k) + \frac{M}{J_n} \sum_{k=1}^{J_n} C_k I(t \geq t_k)$$

Replacing the logarithm of the likelihood function (4) by the piecewise assumption of $H(\cdot)$, the log-likelihood function becomes

$$\frac{1}{n}\sum_{i=1}^{n}(-\delta_i \boldsymbol{\beta}'\mathbf{x}_\mathbf{i}^*) + \frac{1}{n}\sum_{k=1}^{J_n} \log C_k \Big\{ \sum_{i=1}^{n} \delta_i I(e^{r_i(\boldsymbol{\beta})} \in [t_{k-1}, t_k)) \Big\}$$
$$-\frac{1}{n}\sum_{k=1}^{J_n} C_k \Big\{ \sum_{i=1}^{n} (e^{r_i(\boldsymbol{\beta})} - t_k) I(t_{k-1} \leq e^{r_i(\boldsymbol{\beta})} < t_k) + \frac{M}{J_n}\sum_{i=1}^{n} I(e^{r_i(\boldsymbol{\beta})} \geq t_k) \Big\} \quad (5)$$

After the calculation (Zeng and Lin (2007)), the kernel-smoothed approximation of the log-likelihood function is

$$l_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i \{ -\boldsymbol{\beta}\mathbf{x}_i^* - r_i(\boldsymbol{\beta}) + \log[\frac{1}{na_n}\sum_{j=1}^{n} \delta_j K(\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{a_n})]$$
$$- \log[\frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\frac{r_j(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})}{a_n}} K(s)ds] \} \quad (6)$$

where a_n is the bandwidth, and $\boldsymbol{K}(\cdot)$ is the kernel-smoothed function.

Zeng and Lin (2007) used four bandwidths, which were $\sigma n^{-1/5}$, $\sigma n^{-1/7}$, $\sigma n^{-1/9}$ and optimal bandwidth, where σ was estimated from the standard deviation of $\log(T)$. The optimal bandwidth includes two parts: $(8\sqrt{2}/3)^{1/5}\sigma_1 n^{-1/5}$ and $4^{1/3}\sigma_2 n^{-1/3}$ (Jones (1990), Jones and Sheather (1991)), where σ_1 was estimated from the standard deviation of $\log(T) - \boldsymbol{\beta}'\mathbf{x}^*$ among uncensored subjects and σ_2 was estimated from the standard deviation of $\log(T) - \boldsymbol{\beta}'\mathbf{x}^*$ among all subjects. The bandwidth $(8\sqrt{2}/3)^{1/5}\sigma_1 n^{-1/5}$ is used in the kernel density, and the bandwidth $4^{1/3}\sigma_2 n^{-1/3}$ is used in the cumulative kernel density. The normal density is used for the kernel-smoothed function.

The estimation of $\boldsymbol{\beta}$ can be obtained by maximizing (6) through the Newton Raphson algorithm. The variance of $\hat{\boldsymbol{\beta}}$ can be estimated by the inverse of negative second derivative of $l_n(\boldsymbol{\beta})$, which is

$$Var(\hat{\boldsymbol{\beta}}) \approx (-\frac{\partial^2 l_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2})^{-1}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \quad (7)$$

Based on $\hat{\boldsymbol{\beta}}$, the hazard function $h_T(t)$ can be estimated by

$$\hat{h}_T(t) = \frac{\frac{1}{na_n t}\sum_{i=1}^{n} \delta_i K(\frac{r_i(\hat{\boldsymbol{\beta}}) - log(t)}{a_n})}{\frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\frac{r_i(\hat{\boldsymbol{\beta}}) - log(t)}{a_n}} K(u)du} \quad (8)$$

The cumulative hazard function $H_T(t)$ can be estimated by

$$\hat{H}_T(t) = \int_{-\infty}^{logt} \frac{\frac{1}{na_n}\sum_{i=1}^{n}\delta_i K(\frac{r_i(\hat{\boldsymbol{\beta}})-s}{a_n})}{\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{r_i(\hat{\boldsymbol{\beta}})-s}{a_n}} K(u)du} ds \tag{9}$$

and the corresponding survival function is

$$\hat{S}_T(t) = e^{-\hat{H}_T(t)} \tag{10}$$

3. Simulation Study

This simulation is designed to investigate the impact of sample sizes, bandwidths, and censoring rates on the estimation of regression parameters and the survival function.

In our study, we will use the same model employed by Zeng and Lin (2007):

$$\log T = 2 + X_1 + X_2 + \epsilon \tag{11}$$

where X_1 is a binary variable taking 0 or 1 with the equal probability, and X_2 is drawn from a standard normal distribution with the mean 0 and the standard deviation 1. X_1 and X_2 are independent.

The error term can be generated from a standard normal distribution or a standard extreme value distribution. Censoring time can be generated from the uniform distribution $U(0, \eta)$, and η is used to control the censoring rates. For example, we choose η as 200 to generate 15% censoring rate under the standard normal distribution. In order to investigate the performances under different sample sizes and censoring rates, we set the sample sizes as 200, 400, and 800, and the censoring rates as 15%, 30%, and 45%. 1000 dataset are generated under each situation and all the simulations are conducted in R 2.11.1.

The results of the bias, the estimated standard error (SE), the estimated standard deviation (SD) and the coverage probability (CP) of both $\hat{\beta}_1$ and $\hat{\beta}_2$, and the results of the absolute value of bias (Abs(bias)) between the estimated survival probability and the theoretical survival probability are shown in Tables 1-2.

3.1 Influence of Sample Sizes

From Tables 1-2, we can see that the biases of $\hat{\beta}_1$ always regularly change when the sample size changes from 200 to 800. That is, the bias of $\hat{\beta}_1$ always decreases along with the increase of the sample size, no matter the censoring rates are lower or higher, as well as the error distributions are standard normal error distributions or standard extreme-value error distributions. For example, in Table 1, when the censoring rate is 30%, the bandwidth is σ $n^{-\frac{1}{5}}$, and the sample sizes are (200, 400, 800), the corresponding biases of $\hat{\beta}_1$ are (.0235, .0145, .0091). In Table 2, when the censoring rate is 15%, the bandwidth is $\sigma\ n^{-\frac{1}{7}}$, and the sample sizes are (200, 400, 800), the corresponding biases of $\hat{\beta}_1$ are (.0367, .0264, .0254).

The similar trends of the biases of $\hat{\beta}_2$ can be found from Tables 1-2. That is, when the sample size increases from 200 to 800, the bias of $\hat{\beta}_2$ always decreases, no matter what the error distribution and the bandwidth are. For example, in Table 1, when the censoring rate is 15%, the bandwidths are $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and $optimal$, the corresponding biases of $\hat{\beta}_2$ are (.0124, .0081, .0079), (.0166, .0110, .0105), (.0197, .0137, .0133), and (.0143, .0079, .0068) for sample size (200, 400, 800).

Under the standard normal error distribution, when the bandwidths are $\sigma\ n^{-\frac{1}{7}}$ and σ $n^{-\frac{1}{9}}$, the bias of the survival probability always decreases while the sample size increases

Table 1: Bias, SE, SD, and coverage probability of $\hat{\beta}$ from the 1000 dataset with standard normal error distribution

			$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{S}_T(t)$
Censoring rate (%)	n	Bandwidth (a_n)	Bias	SE	SD	CP(%)	Bias	SE	SD	CP(%)	Abs(bias)
15	200	$\sigma\, n^{-\frac{1}{5}}$	.0210	.1547	.1589	93.5	.0124	.0797	.0827	93.0	.0060
		$\sigma\, n^{-\frac{1}{7}}$	.0233	.1599	.1545	94.7	.0166	.0826	.0795	95.4	.0082
		$\sigma\, n^{-\frac{1}{9}}$	.0254	.1641	.1545	95.1	.0197	.0848	.0792	96.0	.0118
		$optimal$	.0216	.1539	.1545	94.0	.0143	.0793	.0795	94.3	.0075
	400	$\sigma\, n^{-\frac{1}{5}}$	.0080	.1080	.1125	93.4	.0081	.0556	.0577	93.8	.0059
		$\sigma\, n^{-\frac{1}{7}}$	.0123	.1113	.1094	95.0	.0110	.0574	.0561	95.7	.0067
		$\sigma\, n^{-\frac{1}{9}}$	.0154	.1142	.1098	95.6	.0137	.0588	.0564	96.2	.0096
		$optimal$	.0087	.1072	.1085	94.5	.0079	.0554	.0555	95.0	.0090
	800	$\sigma\, n^{-\frac{1}{5}}$	.0047	.0756	.0769	94.0	.0079	.0389	.0418	92.9	.0083
		$\sigma\, n^{-\frac{1}{7}}$	.0073	.0776	.0751	94.5	.0105	.0398	.0407	94.1	.0063
		$\sigma\, n^{-\frac{1}{9}}$	.0100	.0795	.0751	95.0	.0133	.0408	.0407	93.6	.0077
		$optimal$	.0039	.0749	.0754	93.8	.0068	.0386	.0407	93.6	.0113
30	200	$\sigma\, n^{-\frac{1}{5}}$	.0235	.1652	.1770	92.3	.0224	.0880	.0943	91.8	.0067
		$\sigma\, n^{-\frac{1}{7}}$	.0326	.1713	.1727	94.5	.0295	.0912	.0925	94.6	.0098
		$\sigma\, n^{-\frac{1}{9}}$	.0401	.1764	.1735	94.8	.0364	.0939	.0932	94.8	.0147
		$optimal$	.0234	.1638	.1719	93.7	.0212	.0872	.0915	93.5	.0084
	400	$\sigma\, n^{-\frac{1}{5}}$	.0145	.1143	.1197	94.3	.0187	.0607	.0635	91.4	.0065
		$\sigma\, n^{-\frac{1}{7}}$	.0214	.1183	.1175	94.8	.0256	.0626	.0628	93.2	.0080
		$\sigma\, n^{-\frac{1}{9}}$	.0277	.1219	.1187	95.3	.0321	.0644	.0636	93.2	.0124
		$optimal$	.0122	.1130	.1164	94.6	.0164	.0599	.0622	92.6	.0099
	800	$\sigma\, n^{-\frac{1}{5}}$	.0091	.0803	.0821	94.0	.0122	.0425	.0411	93.8	.0070
		$\sigma\, n^{-\frac{1}{7}}$	.0168	.0827	.0803	94.9	.0199	.0437	.0405	93.9	.0063
		$\sigma\, n^{-\frac{1}{9}}$	.0236	.0851	.0808	95.1	.0268	.0449	.0411	92.6	.0099
		$optimal$	.0065	.0793	.0807	94.3	.0096	.0420	.0403	94.9	.0112
45	200	$\sigma\, n^{-\frac{1}{5}}$	.0402	.1823	.1883	94.2	.0359	.0993	.1056	91.3	.0088
		$\sigma\, n^{-\frac{1}{7}}$	.0530	.1907	.1849	95.7	.0500	.1038	.1056	91.9	.0158
		$\sigma\, n^{-\frac{1}{9}}$	.0651	.1976	.1875	95.6	.0629	.1078	.1076	92.1	.0240
		$optimal$	.0362	.1801	.1810	95.1	.0332	.0978	.1021	92.4	.0084
	400	$\sigma\, n^{-\frac{1}{5}}$	.0278	.1262	.1264	94.5	.0241	.0689	.0715	93.4	.0065
		$\sigma\, n^{-\frac{1}{7}}$	.0426	.1311	.1250	94.9	.0386	.0717	.0706	92.8	.0117
		$\sigma\, n^{-\frac{1}{9}}$	.0560	.1359	.1267	95.5	.0519	.0743	.0719	91.4	.0191
		$optimal$	.0231	.1242	.1226	94.7	.0191	.0679	.0687	94.4	.0106
	800	$\sigma\, n^{-\frac{1}{5}}$	.0211	.0876	.0894	93.9	.0190	.0477	.0505	91.6	.0060
		$\sigma\, n^{-\frac{1}{7}}$	.0347	.0908	.0892	93.8	.0327	.0494	.0504	89.5	.0088
		$\sigma\, n^{-\frac{1}{9}}$	.0476	.0941	.0908	93.4	.0456	.0512	.0515	86.3	.0156
		$optimal$	.0149	.0862	.0883	94.0	.0131	.0470	.0496	92.8	.0104

from 200 to 800. For example, in Table 1, when the censoring rate is 30%, the bandwidths are $\sigma\ n^{-\frac{1}{7}}$ and $\sigma\ n^{-\frac{1}{9}}$, the corresponding biases of $\hat{S}_T(t)$ are (.0098, .0080, .0063) and (.0147, .0124, .0099). However, when the bandwidths are $\sigma\ n^{-\frac{1}{5}}$ and $optimal$, the increase of sample sizes does not always lead to the decrease of bias of $\hat{S}_T(t)$, which may be better if the sample size is larger than 800. Under the standard extreme-value error distribution, when the censoring rates are 15% and 30%, the bias of the survival probability always decreases while the sample size increases from 200 to 800. Similar trend can be found in the censoring rate of 45%, except for the bandwidth $optimal$. The possible reason is that the $optimal$ bandwidth may be effected easily when the censoring rate is large, such as 45%.

3.2 Influence of Bandwidths

It can be easily seen that the bias of $\hat{\beta}_1$ has the same trend: bias of $\hat{\beta}_1$ increases with the bandwidth changes among $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$, and then decreases with the bandwidth changes from $\sigma\ n^{-\frac{1}{9}}$ to $optimal$. For instance, in Table 1, when the censoring rate is 45%, the sample size is 400, and the bandwidths are $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and $optimal$, the corresponding biases of $\hat{\beta}_1$ are (.0278, .0426, .0560, .0231). Similar findings for the influence of the bandwidths on the biases of $\hat{\beta}_2$ can be seen from Tables 1-2. For example, in Table 2, when the censoring rate is 45%, the sample size is 200, and the bandwidths are $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and $optimal$, the corresponding biases of $\hat{\beta}_2$ are (.0747, .0995, .1178, .0718).

Under the standard normal error distribution, when the sample sizes are 200 and 400, the bias of $\hat{S}_T(t)$ firstly increases among the three bandwidths: $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and then decreases between the two bandwidths: $\sigma\ n^{-\frac{1}{9}}$ and $optimal$. However, when the sample size is 800, the bias of $\hat{S}_T(t)$ does not always increase among the previous three bandwidths, which again shows that when the sample size is larger than 800, the bias of $\hat{S}_T(t)$ may have the similar trends like those of sample size 200 and 400. Under the standard extreme-value error distribution, similarly, the bias of $\hat{S}_T(t)$ firstly increases among the previous there bandwidths and then decreases between the last two bandwidths, no matter what sample sizes are. For instance, when the censoring rate is 30%, and the sample sizes are 200, 400 and 800, the corresponding biases of $\hat{S}_T(t)$ are (.0312, .0428, .0531, .0241), (.0268, .0374, .0471, .0208) and (.0045, .0081, .0136, .0047), for the bandwidths $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and $optimal$.

3.3 Influence of Censoring Rate

It is obvious that along with the censoring rate increasing from 15% to 45%, the biases of $\hat{\beta}_1$ and $\hat{\beta}_2$ increase. For example, in Table 1, we can see that when the sample size is 200 and the bandwidth is $\sigma\ n^{-\frac{1}{9}}$, the corresponding biases of $\hat{\beta}_1$ for 15%, 30% and 45% are (.0254, .0401, .0651), and the corresponding biases of $\hat{\beta}_2$ for 15%, 30% and 45% are (.0197, .0364, .0629).

Under the standard normal error distribution, when the bandwidths are $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$, the bias of $\hat{S}_T(t)$ increases along with the censoring rate increases from 15% to 45%. For the other two bandwidths, the change pattern of the bias of $\hat{S}_T(t)$ is not obvious. Under the standard extreme-value error distribution, the bias of $\hat{S}_T(t)$ does not always increase along with the censoring rate increases from 15% to 45%. These results show that the uniform distribution where the censoring time is generated may have some influence on the change pattern of the bias of $\hat{S}_T(t)$. Thus, censoring time generated from other type of distributions may cause the obvious change pattern of the bias of $\hat{S}_T(t)$.

Table 2: Bias, SE, SD, and coverage probability of $\hat{\beta}$ from the 1000 dataset with standard extreme-value error distribution

			$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{S}_T(t)$
Censoring rate (%)	n	Bandwidth (a_n)	Bias	SE	SD	CP(%)	Bias	SE	SD	CP(%)	Abs(bias)
15	200	$\sigma\, n^{-\frac{1}{5}}$	.0297	.1821	.1870	93.3	.0250	.0944	.0993	92.6	.0263
		$\sigma\, n^{-\frac{1}{7}}$	.0367	.1955	.1877	95.2	.0332	.1018	.1013	93.6	.0349
		$\sigma\, n^{-\frac{1}{9}}$	.0414	.2053	.1911	95.7	.0388	.1068	.1041	94.3	.0422
		$optimal$	.0349	.1791	.1824	93.9	.0294	.0927	.0951	93.2	.0218
	400	$\sigma\, n^{-\frac{1}{5}}$	.0170	.1250	.1231	94.4	.0189	.0652	.0663	92.7	.0232
		$\sigma\, n^{-\frac{1}{7}}$	.0264	.1351	.1263	96.0	.0280	.0702	.0682	93.8	.0307
		$\sigma\, n^{-\frac{1}{9}}$	.0324	.1424	.1302	96.1	.0339	.0738	.0701	94.2	.0377
		$optimal$	.0204	.1219	.1194	94.7	.0211	.0635	.0644	93.1	.0194
	800	$\sigma\, n^{-\frac{1}{5}}$	.0167	.0863	.0858	94.5	.0133	.0452	.0462	93.9	.0048
		$\sigma\, n^{-\frac{1}{7}}$	.0254	.0931	.0886	94.8	.0222	.0484	.0471	93.1	.0068
		$\sigma\, n^{-\frac{1}{9}}$	.0309	.0982	.0916	95.4	.0282	.0509	.0488	92.4	.0104
		$optimal$	.0200	.0848	.0833	94.5	.0158	.0443	.0441	93.9	.0052
30	200	$\sigma\, n^{-\frac{1}{5}}$	.0665	.2032	.2204	91.3	.0457	.1080	.1180	90.6	.0312
		$\sigma\, n^{-\frac{1}{7}}$	.0829	.2187	.2196	93.7	.0635	.1166	.1177	92.7	.0428
		$\sigma\, n^{-\frac{1}{9}}$	.0942	.2298	.2243	93.9	.0759	.1225	.1204	92.0	.0531
		$optimal$	.0672	.1983	.2082	92.4	.0484	.1055	.1107	92.0	.0241
	400	$\sigma\, n^{-\frac{1}{5}}$	.0402	.1396	.1437	93.1	.0366	.0750	.0800	91.1	.0268
		$\sigma\, n^{-\frac{1}{7}}$	.0574	.1505	.1439	94.0	.0542	.0807	.0804	90.6	.0374
		$\sigma\, n^{-\frac{1}{9}}$	.0693	.1586	.1480	93.6	.0666	.0847	.0829	89.1	.0471
		$optimal$	.0381	.1354	.1369	92.8	.0351	.0728	.0748	91.8	.0208
	800	$\sigma\, n^{-\frac{1}{5}}$	.0328	.0962	.1020	92.2	.0293	.0522	.0549	91.2	.0045
		$\sigma\, n^{-\frac{1}{7}}$	.0520	.1037	.1034	92.5	.0481	.0558	.0563	86.6	.0081
		$\sigma\, n^{-\frac{1}{9}}$	.0657	.1096	.1067	92.1	.0616	.0586	.0585	82.8	.0136
		$optimal$	.0323	.0941	.0967	93.3	.0301	.0509	.0521	91.8	.0047
45	200	$\sigma\, n^{-\frac{1}{5}}$	.0843	.2298	.2652	87.5	.0747	.1250	.1433	87.5	.0101
		$\sigma\, n^{-\frac{1}{7}}$	.1107	.2510	.2622	90.3	.0995	.1366	.1439	89.4	.0204
		$\sigma\, n^{-\frac{1}{9}}$	.1305	.2646	.2661	91.9	.1178	.1436	.1480	88.7	.0306
		$optimal$	.0796	.2265	.2495	89.3	.0718	.1233	.1345	90.4	.0023
	400	$\sigma\, n^{-\frac{1}{5}}$	.0637	.1594	.1750	91.0	.0659	.0874	.0980	86.5	.0068
		$\sigma\, n^{-\frac{1}{7}}$	.0937	.1721	.1758	90.8	.0927	.0943	.0971	85.0	.0152
		$\sigma\, n^{-\frac{1}{9}}$	.1154	.1813	.1809	90.4	.1131	.0993	.0998	81.5	.0246
		$optimal$	.0556	.1541	.1641	92.0	.0584	.0848	.0920	88.5	.0028
	800	$\sigma\, n^{-\frac{1}{5}}$	.0420	.1098	.1142	91.8	.0454	.0604	.0612	89.2	.0047
		$\sigma\, n^{-\frac{1}{7}}$	.0704	.1181	.1169	91.4	.0739	.0647	.0622	81.7	.0111
		$\sigma\, n^{-\frac{1}{9}}$	.0920	.1247	.1216	89.5	.0956	.0681	.0646	73.6	.0192
		$optimal$	.0368	.1065	.1073	93.3	.0407	.0588	.0577	90.6	.0041

4. Real Data Analysis

For illustration, we apply the profile likelihood method to the Methadone Maintenance Treatment study. The methadone maintenance treatment is one type of therapies used to help people recover from heroin dependence, and help them to live a productive life. This dataset (Caplehorn and Bell (1991)) has 238 heroin addicts, who took part in the maintenance programme from February 1986 to August 1987. The main interest of this study is to investigate whether the clinic dosage policies have the significant effect on retaining addicts in methadone maintenance treatment.

Addicts entering into this study were assigned into clinic 1 ($clinic$ = 1) or clinic 2 ($clinic$ = 2) based on the dosage policies. Two other covariates ($prison$ and $dose$) were also investigated by Caplehorn and Bell (1991). The variable $prison$ was coded as “0” when prison records were not present, and “1” when prison records were present. $dose$ was a continuous variable indicating the maximum methadone dose used by addicts per day. The median survival time for this dataset is 512 days with 95% confidence interval (399, 563).

Table 3: Estimates, standard errors (SE), 95% confidence intervals and p-values of parameters from different bandwidths for the methadone maintenance treatment data

bandwidths		clinic	prison	dose
	Estimate	.8347	-.2350	.0300
	SE	.1676	.1367	.0047
$\sigma\, n^{-\frac{1}{5}}$	CI upper	1.163	.0329	.0392
	CI lower	.5062	-.5029	.0208
	p-value	$< .0001$	.0868	$< .0001$
	Estimate	.7804	-.2249	.0329
	SE	.1936	.1526	.0054
$\sigma\, n^{-\frac{1}{7}}$	CI upper	1.160	.0742	.0435
	CI lower	.4009	-.5240	.0223
	p-value	.0001	.1419	$< .0001$
	Estimate	.7582	-.2417	.0348
	SE	.2133	.1667	.0059
$\sigma\, n^{-\frac{1}{9}}$	CI upper	1.176	.0850	.0464
	CI lower	.3401	-.5684	.0232
	p-value	.0005	.1483	$< .0001$
	Estimate	.7974	-.1961	.0300
	SE	.1623	.1328	.0046
$optimal$	CI upper	1.116	.0642	.0390
	CI lower	.4793	-.4564	.0210
	p-value	$< .0001$	.1410	$< .0001$

In order to quantify the differences between these two groups by the PH model, we need to check the PH assumption. We conduct the Schoenfeld residual test to check the PH assumption by using “cox.zph” in R, which gives us the p-value .0325. In other words, the PH assumption is not satisfied for this dataset. Therefore, methadone maintenance treatment data are fitted by the following AFT model:

$$\log(T) = \beta_1 \times clinic + \beta_2 \times prison + \beta_3 \times dose + \epsilon$$

Table 3 shows the estimate, standard error (SE), 95% confidence interval and p-value of parameters from different bandwidths for the methadone maintenance treatment data. It

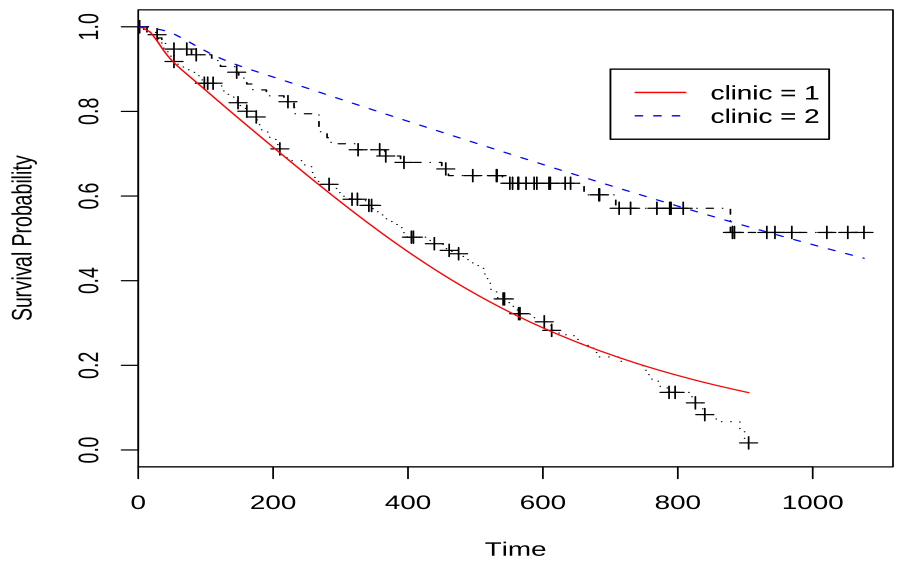

Figure 1: Survival curves from profile likelihood method and Kaplan Meier method for two clinic groups

can be seen that along with the changes of the bandwidths, the estimates change a little. But the changes of bandwidths do not have effects on the significance of $clinic$, $prison$ and $dose$. For example, when the bandwidths change among $\sigma\ n^{-\frac{1}{5}}$, $\sigma\ n^{-\frac{1}{7}}$, $\sigma\ n^{-\frac{1}{9}}$ and $optimal$, the corresponding estimates of $clinic$ are (.8347, .7804, .7582, .7974), and the corresponding p-values for $clinic$ are ($<$ .0001, .0001, .0005, $<$ .0001), which show the significant effect of $clinic$. This trend can also be seen from the 95% confidence intervals: $\sigma\ n^{-\frac{1}{5}}$: (.5062, 1.163), $\sigma\ n^{-\frac{1}{7}}$: (.4009, 1.160), $\sigma\ n^{-\frac{1}{9}}$: (.3401, 1.176), $optimal$: (.4793, 1.116). It is obvious that 0 is not included in any of the four confidence intervals, which indicates that $clinic$ has the significant effect on the survival time.

When the bandwidth is $\sigma\ n^{-\frac{1}{5}}$, the AFT model can be fitted as:

$$\log(T) = .8347 \times clinic - .2350 \times prison + .0300 \times dose$$

Let other covariates be fixed, and the survival time for the clinic 2 is 2.304 times larger than the clinic 1. Given the average values of $prison$ (.46) and $dose$ (59.16) of clinic 1, as well as the average values of $prison$ (.49) and $dose$ (63.43) of clinic 2, the survival curves can be estimated by using the profile likelihood method. Figure 1 displays the fitted survival curves from both the profile likelihood method and the KM method for two clinic groups, which indicates that the estimated survival curves from the profile likelihood method fit the nonparametric estimator well. From these figures, we can see addicts in clinic 2 have a better survival probability than those in clinic 1.

5. Conclusions and Discussions

We applied semi-parametric AFT model based on the profile likelihood method proposed by Zeng and Lin (2007) and investigated the performance of the profile likelihood method by the comprehensive simulation study. The accuracy of estimates were varied by the bandwidths. We recommend the profile likelihood method with $\sigma\ n^{-\frac{1}{5}}$ bandwidth, since this bandwidth has smaller biases of both estimated parameters and estimated survival probabilities.

REFERENCES

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

Tsiatis, A. A. (1990), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *The Annals of Statistics*, 18, 354-372.

Caplehorn, J. R., and Bell, J. (1991), "Methadone Dosage and Retention of Patients in Maintenance Treatment," *The Medical Journal of Australia*, 154, 195-199.

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

Lai, T. L., and Ying, Z. L. (1991), "Rank Regression Methods for Left-Truncated and Right-Censored Data," *The Annals of Statistics*, 19, 531-556.

Stute, W., and Wang, J. L. (1993), "The Strong Law under Random Censorship," *The Annals of Statistics*, 21, 1591-1607.

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.

Huang, J., Ma, S. G., and Xie, H. L. (2006), "Regularized Estimation in the Accelerated Failure Time Model with High-Dimensional Covariates," *Biometrics*, 62, 813-820.

Jin, Z. Z., Lin, D. Y., and Ying, Z. L. (2006), "On least-squares regression with censored data," *Biometrika*, 93, 147-161.

Huang, L., and Jin, Z. Z. (2007), "LSS: An S-Plus/R program for the accelerated failure time model to right censored data based on least-squares principle," *Computer Methods and Programs in Biomedicine*, 86, 45-50.

Zeng, D. L., and Lin, D. Y. (2007), "Efficient Estimation for the Accelerated Failure Time Model," *Journal of the American Statistical Association*, 102, 1387-1396.

Brown, B. M., and Wang, Y. G. (2007), "Induced smoothing for rank regression with censored survival times," *Statistics in Medicine*, 26, 828-836.

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.