Statistics and Its Interface Volume 11 (2018) 531–540

A unified semi-empirical likelihood ratio confidence interval for treatment effects in the two sample problem with length-biased data

Tao Li*, Mengyun Wu, and Yong Zhou

In two sample studies, the treatment effects that we are interested in may have different types, such as mean difference, the difference of probabilities, etc. In this work, we apply semi-parametric empirical likelihood principle to length biased data and derived a unified empirical likelihood ratio confidence interval for treatment effects. The empirical likelihood ratio is shown to be asymptotically distributed as chi-squared. Simulation studies show that the proposed confidence interval has a better performance compared with its counterpart which is based on normal approximation. The severe effect caused by ignoring the length bias is also illustrated by simulation. The proposed method is applied to Oscar data to study the effect of high socio-economic status on lifetime.

Keywords and phrases: Empirical likelihood, Estimating equation, Treatment effect, Censored data, Length-biased.

1. INTRODUCTION

In survival analysis, it's very common and necessary to estimate the effect of a treatment compared with a control group or estimate the difference between two treatments. Although it originates in a medical literature concerned with the causal effects of binary, yes-or-no 'treatments', the term 'treatment effect' is now applied, more generally, to other fields of natural and social science, especially psychology, political science, and economics. For instance, Redelmeier and Singh [19] compiled Oscar data to study the effect of high socio-economic status on health and lifetime: among Hollywood performers, does winning an Oscar Award cause their expected lifetime to increase? Some analysis by [19] was based on mean lifetime differences between the Oscar winners and the performers who were never nominated. In this scenario, treatment means that the performer is nominated for an Oscar Award. Along with the generalization of 'treatment', the measure on its 'effect' associated with the real problem can be presented by different types of statistics. For example, Lin and Zhou [12] also uses Oscar data to study the effect of high socio-economic status on health and lifetime, but their discussion is not only based on the mean lifetime differences but also the probability that the lifetime of an Oscar winner is longer than the performer who is never nominated.

There are a variety of parametric and nonparametric methods proposed to estimate the treatment effects for two-sample problem. In many applications, the historical information is available for the control group but unavailable for the treatment group. In this situation, it is more reasonable to apply the semiparametric model, in which one population is modeled parametrically and the other nonparametrically. Li et al. [11] proposed semiparametric estimate for the quantile comparison function and showed that it could have substantially smaller asymptotic variance than its nonparametric counterpart proposed by [10]. One of its special cases, where one population is assumed to be normally distributed, has been discussed by [9]. A unified semiparametric method for different types of treatment effects was proposed by [24] for right-censored data. Based on the same semiparametric model, another estimate was discussed by [4].

This paper is motivated by discussions about Oscar data [21, 7, 12]. Based on the analysis of mean lifetime differences, it is stated by [19] that life expectancy is 3.9 years longer for Oscar Award winners than other less recognized performers. But later it is pointed out by [21] that the analysis in [19] suffers from immortal time bias—the performers who live longer have more opportunities to be nominated. Consider the example given by [7], both Henry Fonda and Dan Dailey were first nominated for an Oscar Award but did not win at the age of 35. Fonda finally won the Oscar at age 77, while Dan died at age 64 and never won the Oscar before death. Which means, by taking the age of the first time nomination as the truncation time, the observations in the nominee group (no matter if winning the award or not) are actually left truncated. As we know, the left truncated data are very common in prevalent study. Particularly, when the incidence of the event follows a stationary Poisson process, the left truncated data is called 'the length-biased data' [2, 3]. As a special case of left truncation, the length bias means that the left truncation time is uniformly distributed, and the sampling probability of survival time is proportional to its length. With the length-biased sampling

*Corresponding author.

procedure being adopted, it is getting important to develop the methodology for statistical inference of treatment effects in the two-sample problem with the length bias and right censoring. Lin and Zhou [12] pooled all the nominees (including both winners and non-winning nominees) as the treatment group (nominee group), and took the performers who were never nominated as the control (non-nominee group). By taking the age of the first time nomination as the truncation time, they used the method proposed by [1] to check the stationarity for truncation time and showed that the length-bias assumption on lifetime data in the nominee group was applicable. By applying a parametric model to length-biased and right-censored data (nominee group) and a nonparametric model to right-censored data (non-nominee group), they proposed a unified semiparametric estimating equation approach to estimate various types of treatment effects. The large sample properties of the estimates were discussed and the normal approximation confidence intervals were constructed by the asymptotic normal distribution of the point estimator.

In this paper, we will consider the same model as [12]. The contribution of this paper is to construct the confidence intervals for several different types of treatment effect measurements in the two-sample problem with the length bias by using empirical likelihood ratio method. Empirical likelihood ratio method was first introduced by [22], then developed by [13, 14] and many others into a general methodology. Compared with normal approximation method and bootstrap method for constructing the confidence interval, empirical likelihood ratio method, as a data-driven method, has many advantages in its accuracy, implementation and flexibility. As we know, a normal approximation confidence interval constructed by the asymptotic normal distribution of a point estimator [12] has to be symmetric implied by the asymptotic normality, whereas the empirical likelihood confidence interval does not have predetermined shape. Empirical likelihood confidence interval may have a better performance than normal approximation confidence interval in terms of coverage probability and it overcomes the undercoverage problem of the normal approximation confidence interval for small sample size [23]. Compared with the bootstrap method, empirical likelihood can be Bartlett corrected, improving the accuracy of inferences [6]. Moreover, empirical likelihood confidence interval preserves the range of the parameter. Likelihoods also make it easier to combine data from multiple sources, with possibly different sampling schemes. All these appealing features make the empirical likelihood ratio confidence region applicable in a variety of situations; see for example [5, 18, 8]. There has been a rich body of literature on the empirical likelihood method in various settings of statistical inference. We here just concentrate our review on the two-sample problem. Qin [15] combined the empirical likelihood ideas and the parametric method to construct confidence intervals for the difference of two population means in a semiparametric model. Later on, this method was applied to right-censored data by [24, 17]. In this article, we derive the asymptotic chi-squared distribution of the empirical likelihood ratio and construct the confidence region for the treatment effects based on chi-squared distribution. Then simulation studies show that the proposed confidence interval has a better performance than its counterpart which is based on normal approximation. The severe effect caused by ignoring the length bias is also illustrated by simulation. The proposed method is applied to Oscar data to study the effect of high socio-economic status on lifetime.

The rest of article is organized as follows. Section 2 introduces the notation, sets up the model and applies the empirical likelihood principle to the model. Section 3 presents the confidence interval based on empirical likelihood principle. At last, the numerical studies which include both simulation study and real data analysis are conducted in Section 4. All detailed proofs are given in Appendix.

2. SEMI-PARAMETRIC EMPIRICAL LIKELIHOOD RATIO

2.1 Model set-up

Let $\tilde{X}$ and $\tilde{Y}$ be two independent nonnegative random variables. Assume random variable $\tilde{X}$ be with the length-biased sampling and random censoring model. Let A denote the left-truncation time, which is assumed to be uniformly distributed and independent with $\tilde{X}$. In the length-biased sampling procedure, $\tilde{X}$ can be observed only when $\tilde{X} \geq A$. Denote by $X^0 = A + T$ the survival time under a length-biased sampling procedure, in which T denotes the residual life time from recruitment. Then (X^0, A) has the same joint distribution as $(\tilde{X}, A)|\tilde{X} \geq A$. Let C be the residual censoring time from recruitment and independent with $(\tilde{X}, A)$. Then only the censored survival time $X = \min(X^0, A+C)$ is observed. The observations $\{(X_i, A_i, \delta_i), i = 1, \cdots, n\}$ are n independent and identically distributed copies of (X, A, δ), where $\delta = I(X^0 \leq A + C)$.

The random censoring model is assumed for random variable $\tilde{Y}$. Let V be the censoring variable, which is independent with $\tilde{Y}$. Define $Y = \min(\tilde{Y}, V)$ and $d = I(\tilde{Y} \leq V)$. The observations $\{(Y_j, d_i), j = 1, \cdots, m\}$ are m independent and identically distributed copies of (Y, d).

Usually, there is enough information available for the control treatment but limited information available for the new treatment group. Hence a semiparametric approach is adopted in this article. A nonparametric distribution is assumed for $\tilde{X}$. Let $F_1(x)$, $S_1(x)$ and $f(x)$ be the distribution function, survival function and density function respectively. A parametric form is assumed for $\tilde{Y}$. Let $f_2(y;\theta)$ be the density function of $\tilde{Y}$, which is of known form with p-dimensional unknown parameter θ. Denote by $F_2(y;\theta)$ and $S_2(y;\theta)$ the corresponding distribution function and survival function respectively. It is assumed that the two samples are independent and $n/m \to \zeta$ as $n, m \to \infty$.

Let Δ be the q-dimensional parameter of interest, which stands for q types of treatment effects. Assume that there is an unbiased estimating function ψ such that

$$E_{F_1}\psi(\tilde{X},\theta_0,\Delta_0)=0, \tag{1}$$

where θ_0 and Δ_0 are the true values of θ and Δ respectively. The unbiased estimating function ψ can be derived according to the type of interested treatment effect. For different type of treatment effect measurement Δ, its corresponding estimating function is given as follows.

i For mean difference, let $\Delta_0 = E\tilde{X} - E\tilde{Y}$ and $\psi(\tilde{X},\theta_0,\Delta_0)=\tilde{X}-\Delta_0-\int_0^{+\infty} y dF_2(y;\theta_0)$;

ii For the difference of probabilities, i.e., $P(\tilde{X}\le t_0)-P(\tilde{Y}\le t_0)$ for a given t_0, let $\Delta_0=F_1(t_0)-F_2(t_0;\theta)$ and $\psi(\tilde{X},\theta_0,\Delta_0)=I(\tilde{X}\le t_0)-F_2(t_0;\theta)-\Delta_0$;

iii For the probability of event $I(\tilde{X}<\tilde{Y})$, let $\Delta_0 = E_{F_1}(1-F_2(\tilde{X};\theta_0))$ and $\psi(\tilde{X},\theta_0,\Delta_0)=1-F_2(\tilde{X},\theta_0)-\Delta_0$.

iv For the value on the receiver operating characteristic curve, let $\Delta_0 = 1-F_1\{F_2^{-1}(1-p;\theta_0)\}$ for a given $p\in(0,1)$ and $\psi(\tilde{X},\theta_0,\Delta_0)=1-I\{\tilde{X}\le F_2^{-1}(1-p;\theta_0)\}-\Delta_0$.

2.2 Semiparametric empirical likelihood ratio

In this article, we apply the empirical likelihood principle to obtain the confidence interval for Δ. It is noted that $\tilde{X}$ in (1) is unobservable in the length-biased sampling procedure. Therefore, we first set up an estimating equation based on the feature of right-censored and length-biased sample.

Under length-biased sampling, the truncation variable A follows a uniform distribution and the joint density function of (X^0, A) is

$$f_{X^0,A}(x,a)=\frac{f_1(x)I(x\ge a)}{\mu}, \tag{2}$$

where $\mu = E(\tilde{X}) = \int_0^{+\infty} x f_1(x)dx$. Denote by $f_0(x)$ and $S_0(x)$ the density function and survival function of X^0 respectively. Then X^0 has a length-biased density function $f_0(x)=\frac{xf_1(x)}{\mu}$. Under the assumption that the censoring variable C is independent with $(\tilde{X},A)$, the probability of observing a failure at time x is

$$\begin{aligned} P(X\in(x,x+dx),\delta=1) &= \frac{f_1(x)}{\mu}dx\int_0^x S_C(z)dz \\ &= \frac{f_1(x)\pi(x)}{\mu}dx, \end{aligned}$$

where $S_C(\cdot)$ is the survival function of C and $\pi(x)=\int_0^x S_C(z)dz$. By some simple calculation, we have

$$\begin{aligned} E\left(\delta\frac{\psi(X,\theta,\Delta)}{\pi(X)}\right) &= \frac{1}{\mu}\int_0^{+\infty}\psi(X,\theta,\Delta)f_1(x)dx \\ &= \frac{1}{\mu}E_{F_1}\psi(\tilde{X},\theta,\Delta)=0. \end{aligned} \tag{3}$$

Therefore, an unbiased estimating function, $\delta\frac{\psi(X,\theta_0,\Delta_0)}{\pi(X)}$, which is based on the observable variable X under length-biased sampling, is constructed. Under the constraining condition (3), the adjusted empirical likelihood ratio approach is used to derive the confidence interval for Δ. Obviously, there is a nuisance parameter $S_C(\cdot)$ in (3). It is natural to replace it with its Kaplan-Meier estimator $\hat{S}_C(t)=\prod_{u\le t}\left(1-\frac{dN_C(u)}{L(u)}\right)$, where $N_C(u)=\sum_{i=1}^n I(X_i-A_i\le u,\delta_i=0)$ and $L(u)=\sum_{i=1}^n I(X_i-A_i\ge u)$.

Let F_p be the distribution function that assigns probabilities p_i at points $\delta_i/\pi(x_i)$. Thus the adjusted semi-parametric empirical likelihood function is

$$L_{adj}(\theta,\mathcal{P})=\prod_{i=1}^n p_i\prod_{j=1}^m f_2(y_j;\theta)^{d_j}S_2(y_j;\theta)^{1-d_j}, \tag{4}$$

where $\mathcal{P}$ denotes the set of all possible values of p_i with restrictions $\sum_{i=1}^n p_i=1$ and $p_i\ge 0$ $(i=1,\cdots,n)$.

Denote by $\hat{\theta}_{\text{MLE}}$ the MLE based on the sample for $\tilde{Y}$, it is easy to show that when $\hat{p}_i=\frac{1}{n}$ and $\hat{\theta}=\hat{\theta}_{\text{MLE}}$, $L_{adj}(\theta,\mathcal{P})$ attains its maximum value $n^{-n}\prod_{j=1}^m f_2(y_j;\hat{\theta}_{\text{MLE}})^{d_j}S_2(y_j;\hat{\theta}_{\text{MLE}})^{1-d_j}$. Let

$$\begin{aligned} R(\theta,\mathcal{P}) &= \prod_{i=1}^n np_i\prod_{j=1}^m f_2(y_j;\theta)^{d_j}S_2(y_j;\theta)^{1-d_j} \\ &\quad \left[\prod_{j=1}^m f_2(y_j;\hat{\theta}_{\text{MLE}})^{d_j}S_2(y_j;\hat{\theta}_{\text{MLE}})^{1-d_j}\right]^{-1}, \end{aligned}$$

then semiparametric empirical likelihood ratio is

$$\begin{aligned} \mathcal{R}(\Delta)=\sup_{\theta,\mathcal{P}}\Bigg\{R(\theta,\mathcal{P})\,\Bigg|\,&\sum_{i=1}^n p_i=1,\ p_i\ge 0, \\ &\sum_{i=1}^n\frac{p_i\delta_i}{\hat{\pi}(x_i)}\psi(x_i,\theta,\Delta)=0\Bigg\}, \end{aligned}$$

where $\hat{\pi}(x_i)$ is the estimate of $\pi(x_i)$ and $\hat{\pi}(x_i)=\int_0^{x_i}\hat{S}_C(t)dt$.

Let

$$\begin{aligned} \mathcal{L}(\theta,\mathcal{P}) &= -\log R(\theta,\mathcal{P}) \\ &= -\sum_{i=1}^n \log np_i-\ell_2(\theta,\boldsymbol{y})+\ell_2(\hat{\theta}_{\text{MLE}},\boldsymbol{y}) \\ &\triangleq \ell_{\text{EL}}(\theta,\mathcal{P})+\ell_2(\hat{\theta}_{\text{MLE}},\boldsymbol{y}) \end{aligned} \tag{5}$$

where $\ell_2(\theta,\boldsymbol{y})=\sum_{j=1}^m\{d_j\log f_2(y_j;\theta)+(1-d_j)\log S_2(y_j;\theta)\}$. Under the restriction $\sum_{i=1}^n\frac{p_i\delta_i}{\hat{\pi}(x_i)}\psi(x_i,\theta,\Delta)=0$, $\mathcal{L}(\theta,\mathcal{P})$ or $\ell_{\text{EL}}(\theta,\mathcal{P})$ was first minimized with respect to $\mathcal{P}$. By lagrangian multiplier method, we have

$$p_i = p_i(\theta) = n^{-1}\left(1 + w(x_i)\lambda^{\mathrm{T}}(\theta)\psi(x_i,\theta,\Delta)\right)^{-1}, \tag{6}$$

where $w(x_i) = \frac{\delta_i}{\hat{\pi}(x_i)}$ and λ is Lagrange operator. With the condition $\sum_{i=1}^{n} \frac{p_i \delta_i}{\hat{\pi}(x_i)}\psi(x_i,\theta,\Delta) = 0$, $\lambda(\theta)$ is given as the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{w(x_i)\psi(x_i,\theta,\Delta)}{1 + w(x_i)\lambda^{\mathrm{T}}(\theta)\psi(x_i,\theta,\Delta)} = 0. \tag{7}$$

By plugging (6) into (5), the profile empirical log-likelihood ratio is

$$\mathcal{L}(\theta,\Delta) = \ell_{\mathrm{EL}}(\theta,\Delta) + \ell_2(\hat{\theta}_{\mathrm{MLE}}), \tag{8}$$

where

$$\ell_{\mathrm{EL}}(\theta,\Delta) = \sum_{i=1}^{n} \log\left\{1 + w(x_i)\lambda^{\mathrm{T}}(\theta)\psi(x_i,\theta,\Delta)\right\} - \ell_2(\theta). \tag{9}$$

By profiling out the nuisance parameter θ, $max_\theta \mathcal{L}(\theta,\Delta)$, as the function of Δ, can be proved to have a chi-squared distribution. Then the empirical likelihood confidence region for Δ can be constructed through chi-squared distribution.

3. CONFIDENCE INTERVAL FOR TREATMENT EFFECT

In this section we will prove that by minimizing (9) with restriction $\sum_{i=1}^{n} \frac{p_i \delta_i}{\hat{\pi}(x_i)}\psi(x_i,\theta,\Delta) = 0$, there exists an estimate $(\hat{\lambda}, \hat{\theta}_{\mathrm{EL}})$ which lies within an $O_p(n^{-\varrho})$, $\left(\frac{1}{3} < \varrho < \frac{1}{2}\right)$ neighborhood of the true value (λ_0, θ_0). The asymptotic normality of the estimate is then derived. By profiling out θ, we will show that $2\mathcal{L}(\hat{\theta}_{\mathrm{EL}},\Delta)$ is asymptotically distributed as chi-squared with an adjustment factor. Then empirical likelihood confidence region for Δ can be constructed through chi-squared distribution.

All the works in this section are based on the following assumptions:

(i) $\tau_1 \le \tau_C$, where $\tau_1 = \sup\{t : S_1(t) > 0\}$ and $\tau_C = \sup\{t : S_C(t) > 0\}$, and

$$\frac{1}{\mu}\int_0^{\tau_1} \frac{\psi^2(u,\theta,\Delta)}{\pi(u)} dF_1(u) < \infty.$$

(ii) $\psi^2(x,\theta,\Delta), \dot{\psi}_\theta(x,\theta,\Delta)$ and $\dot{\psi}_\Delta(x,\theta,\Delta)$ are continuous and bounded by some function $M_1(x)$ in a neighborhood of the true value (θ_0, Δ_0) such that $\frac{1}{\mu}\int_0^{\tau_1} \frac{M_1^3(u)}{\pi^2(u)} dF_1(u) < \infty$, $E_{F_1}\dot{\psi}_\theta(x,\theta,\Delta) \neq 0$, and $E_{F_1}\dot{\psi}_\Delta(x,\theta,\Delta) \neq 0$.

(iii) The density function $f_2(y;\theta)$ is three times differentiable with respect to θ on $A = \{y : f_2(y;\theta) > 0\}$, which is assumed to be independent of θ. There exists a function $M_2(y)$ such that for any $y \in A$ and θ, $E_\theta|M_2(Y)| < \infty$.

(iv) The information matrix $I(\theta)$ with the entries

$$I_{ij} = -\int_0^{\tau_2} \frac{\partial^2 \log f_2(y;\theta)}{\partial\theta_i \partial\theta_j}(1 - S_V(y)) f_2(y;\theta) dy - \int_0^{\tau_2} \frac{\partial^2 \log S_2(y;\theta)}{\partial\theta_i \partial\theta_j}(S_2(y;\theta) dS_V(y),$$

for $i, j = 1, 2, \cdots, p$, which are continuous and positive definite. Here, $\tau_2 = \sup\{t : S_2(t;\theta)S_V(t) > 0\}$ and $S_V(\cdot)$ is the survival function of V.

Theorem 3.1. *Suppose that Assumptions (i)–(iv) hold. Then, as $n \to \infty$, with probability one $\mathcal{L}(\theta,\Delta)$ attains its maximum value at some point $\hat{\theta}_{\mathrm{EL}}$ in the interior of the ball $\|\theta - \theta_0\| < n^{-\varrho}$ $\left(\frac{1}{3} < \varrho < \frac{1}{2}\right)$, and $\hat{\theta}_{\mathrm{EL}}$ and $\hat{\lambda} = \lambda(\hat{\theta}_{\mathrm{EL}})$ satisfy*

$$Q_{1n}(\hat{\lambda}, \hat{\theta}_{\mathrm{EL}}) = Q_{2n}(\hat{\lambda}, \hat{\theta}_{\mathrm{EL}}) = 0,$$

where

$$Q_{1n}(\lambda,\theta) = \frac{1}{n}\frac{\partial \ell_{\mathrm{EL}}(\theta,\Delta)}{\partial\lambda}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{w(x_i)\psi(x_i,\theta,\Delta)}{1 + w(x_i)\lambda^{\mathrm{T}}\psi(x_i,\theta,\Delta)}, \tag{10}$$

$$Q_{2n}(\lambda,\theta) = \frac{1}{n}\frac{\partial \ell_{\mathrm{EL}}(\theta,\Delta)}{\partial\theta}$$

$$= \frac{\lambda}{n}\sum_{i=1}^{n} \frac{w(x_i)\dot{\psi}_\theta(x_i,\theta,\Delta)}{1 + w(x_i)\lambda^{\mathrm{T}}\psi(x_i,\theta,\Delta)} - \frac{1}{n}\frac{\partial \ell_2(\theta)}{\partial\theta}. \tag{11}$$

Theorem 3.2. *Suppose Assumptions (i)–(iv) hold. Then*

$$\sqrt{n}\begin{pmatrix} \hat{\lambda} \\ \hat{\theta}_{\mathrm{EL}} - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, V_1),$$

where $V_1 = (v_{ij})_{2\times 2}$ is with $v_{11} = C_\theta^2 \Gamma_0 + C_\theta^2 \beta_0^{\mathrm{T}} \Sigma_0^{-1}\beta_0$, $v_{12} = v_{21} = C_\theta \beta_0^{\mathrm{T}}(\Sigma + \beta_0\sigma_0^{-2}\beta_0^{\mathrm{T}})^{-1} - C_\theta^2 \Gamma_0 \beta_0^{\mathrm{T}} \Sigma_0^{-1}$, and $v_{22} = C_\theta^2 \Gamma_0 \Sigma_0^{-1}\beta_0\beta_0^{\mathrm{T}}\Sigma_0^{-1} + (\Sigma_0 + \beta_0\sigma_0^{-2}\beta_0^{\mathrm{T}})^{-1}\Sigma_0 (\Sigma_0 + \beta_0\sigma_0^{-2}\beta_0^{\mathrm{T}})^{-1}$, in which $C_\theta = (\sigma_0^2 + \beta_0^{\mathrm{T}}\Sigma_0^{-1}\beta_0)^{-1}$, $\sigma_0^2 = \sigma_0^2(\theta_0,\Delta)$, $\sigma_0^2(\theta,\Delta) = \frac{1}{\mu}E_{F_1}\left(\frac{\psi^2(X_i;\theta,\Delta)}{\pi(X_i)}\right)$, $\beta_0 = \beta_0(\theta_0,\Delta)$, $\beta_0(\theta,\Delta) = \frac{1}{\mu}E_{F_1}\left(\dot{\psi}_\theta(X_i;\theta,\Delta)\right)$, $\Sigma_0 = \zeta^{-1}I(\theta_0)$, $I(\theta)$ is given in Assumption (iv). $\Gamma_0 = \Gamma(\theta_0,\Delta)$, $\Gamma(\theta,\Delta) = E\left(\frac{\delta(X_i)}{\pi(X_i)}\psi(X_i;\theta,\Delta) + \int_0^{\tau_C} \frac{B(s)}{S_C(s)S_T(s)} dM_i^C(s)\right)^2$, in which $M_i^C(t) = I(X_i - A_i \le t, \delta_i = 0) - \int_0^t I(X_i - A_i \ge s) d\Lambda_C(s)$ and $\Lambda_C(\cdot)$ is the cumulative hazard function of the censored time C, $B(s) = E\left\{\frac{\delta(X_i)}{\pi(X_i)}\psi(X_i;\theta,\Delta)h(s,X_i)\right\} = \frac{1}{\mu}E_{F_1}\{\psi(X_i;\theta,\Delta)h(s,X_i)\}$, and $h(s,t) = \frac{I(s\le t)}{\pi(t)}\int_s^t S_C(u)du$.

Theorem 3.3. *Under Assumptions (i)–(iv), the profile empirical log-likelihood ratio function for Δ, $\mathcal{L}(\hat{\theta}_{\mathrm{EL}},\Delta)$, is such that*

$$2\rho(\theta,\Delta)\mathcal{L}(\hat{\theta}_{\mathrm{EL}},\Delta) \xrightarrow{\mathcal{D}} \chi^2(1),$$

Table 1. Comparison of confidence intervals between semiparametric empirical likehood (SEL) method and semiparametric estimating equation (SEE) method

		$\Delta_1 = E(\tilde{X} - \tilde{Y})$						$\Delta_2 = P(\tilde{X} > \tilde{Y})$					
		SEE			SEL			SEE			SEL		
CR	(n, m)	MP	CP	Length	MP	CP	Length	MP	CP	Length	MP	CP	Length
20%	(50,50)	1.10	93.0	2.67	1.16	93.0	2.47	0.65	91.6	0.27	0.66	91.2	0.23
	(100,100)	1.03	94.8	2.05	1.08	95.2	1.65	0.65	92.6	0.21	0.65	94.6	0.18
	(200,200)	1.08	94.9	1.43	1.13	94.6	1.21	0.65	95.2	0.16	0.65	95.8	0.14
40%	(50,50)	0.99	94.8	2.95	0.93	95.2	2.59	0.64	91.2	0.29	0.64	94.8	0.26
	(100,100)	1.02	95.6	2.07	1.01	96.2	1.67	0.64	94.6	0.23	0.65	95.1	0.20
	(200,200)	1.00	94.9	1.55	0.98	95.1	1.32	0.64	93.6	0.16	0.64	96.4	0.15

'MP' means middle point of the confidence interval. 'CP' means coverage probability. 'Length' means the average length of the confidence interval.

where $\rho(\theta, \Delta) = \frac{\Gamma_0 + \beta_0^{\mathrm{T}} \Sigma_0^{-1} \beta_0}{\sigma_0^2 + \beta_0^{\mathrm{T}} \Sigma_0^{-1} \beta_0}$, *and* $\Gamma_0, \beta_0, \sigma_0^2$ *and* Σ_0 *are defined in Theorem 3.2.*

It should be noted that the adjustment factor $\rho(\theta, \Delta)$ in Theorem 3.3 involves unknown parameters θ and Δ. Therefore, in order to construct the confidence interval for Δ, $\rho(\theta, \Delta)$ is replaced with its estimate $\hat{\rho}(\hat{\theta}_{\mathrm{MLE}}, \hat{\Delta})$, where $\hat{\Delta}$ can be any consistent estimate of Δ. For example, we can choose $\hat{\Delta}$ of [12]. Since both $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\Delta}$ are consistent, $\hat{\rho}(\hat{\theta}_{\mathrm{MLE}}, \hat{\Delta})$ is obviously consistent.

Corollary 3.1. *Suppose that the assumptions of Theorem 3.3 hold, then*

$$\lim_{n \to \infty} P(\Delta \in I_\alpha) = 1 - \alpha,$$

where $I_\alpha = \{\Delta : \ 2\hat{\rho}(\hat{\theta}_{\mathrm{MLE}}, \hat{\Delta}) \mathcal{L}(\hat{\theta}_{\mathrm{EL}}, \Delta) \leq \chi^2_{1-\alpha}(1)\}$ *and* $\chi^2_{1-\alpha}(1)$ *is the* $1 - \alpha$ *quantile of* $\chi^2(1)$.

4. NUMERICAL STUDIES

4.1 Simulation studies

In this section, we conduct simulation studies to evaluate the coverage probability and the average length of 95% semiparametric empirical likelihood (SEL) confidence interval proposed in this article, compare it with semiparametric estimating equation (SEE) normal approximation confidence interval proposed by [12] and show that the proposed approach outperforms the normal approximation method on the basis of shortness of length of confidence interval. Moreover, the severe consequence of ignoring length bias is also assessed by simulation study 2. It shows that ignoring the length bias leads to confidence intervals for parameters that are far beyond the true values and the coverage probabilities are very poor.

Let $\tilde{X}$ be distributed as $\Gamma(\alpha, \beta)$ with $\alpha = \beta = 2$. It can be easily derived that X^0 is distributed as $\Gamma(\alpha + 1, \beta)$. Let the conditional distribution of left truncation variable A conditioned on X^0 be $U(0, X^0)$. Moreover, let the distribution of censoring random variable C be distributed as $Exp(c_1)$ with mean c_1, where the value of c_1 is set to control the censoring rate. Let $\tilde{Y}$ be from $Exp(\theta)$ with mean $\theta = 3$. The censoring variable V is from $Exp(c_2)$, where the value of c_2 is also set to control the censoring rate of observations for $\tilde{Y}$.

Two inferences here are of interest to us. One is the mean difference $\Delta_1 = E(\tilde{X} - \tilde{Y})$, and the other one is the probability of event $I(\tilde{X} > \tilde{Y})$, i.e., $\Delta_2 = P(\tilde{X} > \tilde{Y})$. The corresponding estimating functions can be constructed as

$$\psi_1(\tilde{X}, \theta, \Delta_1) = \tilde{X} - \Delta_1 - \theta, \ \psi_2(\tilde{X}, \theta, \Delta_2) = F_2(\tilde{X}, \theta) - \Delta_2.$$

With all the above settings, we have that the true values of Δ_1 and Δ_2 are 1 and 0.64, respectively. We generate 500 samples, of size $n = m =$ 50,100 and 200 and each with censoring rate 20% and 40%.

Simulation study 1 Lin and Zhou [12] applied semiparametric estimating function method to obtain the point estimate and construct the normal approximation confidence interval. For convenience, their approach is denoted as SEE. Simulation study 1 is designed to evaluate the coverage probability and the average length of 95% confidence interval based on the proposed SEL method and compare it with its competitor SEE. The results are summarized in Table 1. As expected, the length of confidence intervals based on both SEL and SEE decrease along with the increasement of the sample size n and m. When the censoring rate increases from 20% to 40%, the length of confidence intervals based on both SEL and SEE increase. Moreover, it is clear that the confidence intervals based on both SEL and SEE have good coverage probabilities, which are all closed to the nominal level 0.95. Meanwhile, the length of confidence intervals based on SEL are better than SEE in all cases.

Simulation study 2 Simulation study 2 is designed to assess the severe consequence of ignoring length bias. In this study, we ignore the length bias and treat the observations for $\tilde{X}$ as $\Gamma(2, 2)$ right-censored data. Then Kaplan-Meier estimator of survival function of censoring variable C is taken as the the estimate of $\pi(x_i)$. The superscript $\tilde{}$

Table 2. SEL confidence intervals with and without length bias taken into account

		$\Delta_1 = E(\tilde{X} - \tilde{Y})$						$\Delta_2 = P(\tilde{X} > \tilde{Y})$					
		SEL_RC			SEL_ LBRC			SEL_RC			SEL_LBRC		
CR	(n, m)	MP	CP	Length	MP	CP	Length	MP	CP	Length	MP	CP	Length
20%	(50,50)	2.60	25.5	2.06	1.16	93.0	2.47	0.75	28.7	0.17	0.66	91.2	0.23
	(100,100)	1.92	61.2	2.22	1.08	95.2	1.65	0.70	46.4	0.19	0.65	94.6	0.18
	(200,200)	2.55	15.4	1.45	1.13	94.6	1.21	0.73	22.7	0.12	0.65	95.8	0.14
40%	(50,50)	1.74	29.7	1.82	0.93	95.2	2.59	0.69	30.8	0.14	0.64	94.8	0.26
	(100,100)	2.35	4.8	1.00	1.01	96.2	1.67	0.74	17.2	0.09	0.65	95.1	0.20
	(200,200)	1.66	10.6	0.87	0.98	95.1	1.32	0.69	21.1	0.10	0.64	96.4	0.15

'MP' means middle point of the confidence interval. 'CP' means coverage probability. 'Length' means the average length of the confidence interval.

is used to denote the estimates without considering length bias. The resulting estimate of $\pi(x_i)$ is $\tilde{\pi}(x_i) = \tilde{S}_C(x) = \prod_{u \le t}(1 - \frac{d\tilde{N}_C(u)}{\tilde{L}(u)})$, where $\tilde{N}_C(u) = \sum_{i=1}^n I(X_i \le u, \delta_i = 0)$ and $\tilde{L}(u) = \sum_{i=1}^n I(X_i \ge u)$. The corresponding SEL confidence interval for two sample treatment effect based on right-censored data were discussed by [24]. Following their result, the SEL confidence interval without considering length bias is constructed and summarized in Table 2. To avoid confusion, denote by SEL_LBRC and SEL_RC the SEL confidence intervals with and without length bias taken into account respectively. We can see that when length bias is ignored, the average length of the confidence interval has no big difference compared with SEL_LBRC. But the midpoints of confidence interval for both Δ_1 and Δ_2 are far beyond the true values. As a result, the coverage probabilities are very poor. This is not surprising, because the observations for $\tilde{X}$ is actually length biased, which means the units with longer lifetime have more opportunities to be sampled. Therefore, failing to take length bias of $\tilde{X}$ into account will result in the positive bias for estimates of Δ_1 and Δ_2.

4.2 Real data analysis

In this section, we apply the proposed method to Oscar data to study the effect of high socio-economic status on lifetime. Besides, for the purpose of comparison, the SEE method and the SEL method without length bias taken into account (SEL_RC) are also applied. Lin and Zhou [12] pooled all the nominees (including both winners and non-winning nominees) as the treatment group (nominee group), and took the performers who were never nominated as the control (non-nominee group). By considering the age of first time nomination as the truncation time, [12] used the method proposed by [1] to check the stationarity for truncation time and showed that the length-bias assumption on lifetime data in the nominee group was applicable. In this article, we follow their discussion and use our proposed method to derive the confidence intervals for the inferences that we are interested in.

The Oscar data includes the records of 1,670 performers from 1928 to 2000. Due to the wrong information, two cases (ID number 1075 and 1430) are excluded from our computation. Among 1,668 performers, 766 observations with 57.31% censored belonged to nominee group and 902 observations with 48.89% censored belonged to non-nominee group. Let $\tilde{X}$ and $\tilde{Y}$ denote the lifetime for nominee group and non-nominee group respectively. In order to study if being nominated for an Oscar Award would have an impact on the lifetime, two inferences $\Delta_1 = E(\tilde{X} - \tilde{Y})$ and $\Delta_2 = P(\tilde{X} < \tilde{Y})$ are employed. The age of the first time nomination A was taken as the truncation variable. The applicability of the length-bias assumption on lifetime data in nominee group was shown by [12]. To determine the distribution for non-nominee group, [12] obtained the Kaplan-Meier estimate for non-nominee's lifetime and used the Q-Q plot to show that the Weibull distribution $W(\alpha, \beta)$ with $\alpha = 81.3, \beta = 6.8$ is appropriate. We follow the above settings and obtain the maximum likelihood estimates $\hat{\alpha}_{\text{MLE}} = 81.2939$ and $\hat{\beta}_{\text{MLE}} = 6.7887$.

Table 3 presents the confidence intervals for Δ_1 and Δ_2 for Oscar data set based on semiparametric empirical likehood (SEL) method, semiparametric estimating equation (SEE) method and SEL method without length bias taken into account (SEL_RC). Our results are consistent with the results from [12]. All 95% confidence intervals for Δ_1, even the results from SEL_RC, are less than zero, which means that the Oscar Award nomination has the negative impact on lifetime extension. Similarly, all 95% confidence intervals for Δ_2 are greater than 0.5, which also supports the statement that the Oscar Award nomination has the negative impact on lifetime extension. This negative impact may be due to the high pressure or hard work for keeping their high socio-economic level. On the other side, compared with the confidence intervals by SEE and SEL, the confidence interval for Δ_1 by SEL_RC is much closer to zero and the confidence interval for Δ_2 by SEL_RC is closer to 0.5. This phenomenon is quite similar to the simulation study and it is caused by ignoring the length bias for nominee group $\tilde{X}$. Moreover, the confidence intervals based on the proposed method are shorter than the ones based on SEE method for both Δ_1 and Δ_2. This is also similar to the simulation study.

Table 3. Confidence intervals for Oscar data set based on SEL method, SEE method and SEL_RC

	$\Delta_1 = E(\tilde{X} - \tilde{Y})$		$\Delta_2 = P(\tilde{X} < \tilde{Y})$	
	95% CI	Length	95% CI	Length
SEE	(−7.4792, −3.3624)	4.1168	(0.5560, 0.6328)	0.0772
SEL	(−6.4901, −4.4762)	2.0139	(0.5670,0.6170)	0.0500
SEL_RC	(−3.9532, −2.0574)	1.8959	(0.5136,0.5615)	0.0479

APPENDIX A

To prove Theorem 3.1, we first present and prove the following lemmas. The asymptotic normality of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(x_i)}\psi(x_i;\theta,\Delta)$ was proved by [12]. For convenience, we present it as Lemma A.1.

Lemma A.1. *Under the Assumptions (i) and (ii),*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(x_i)}\psi(x_i;\theta,\Delta) \quad \xrightarrow{\mathcal{D}} \quad N(0,\Gamma(\theta,\Delta)),$$

where $\Gamma(\theta,\Delta)$ is defined in Theorem 3.2.

Proof. For a detailed proof please refer to [12]. □

Lemma A.2. *Under the Assumptions (i) and (ii), the following equalities hold.*

$$(i) \qquad \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 = \sigma_0^2(\theta,\Delta)+o_p(1),$$

$$(ii) \qquad \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\dot{\psi}_{\Delta}(x_i;\theta,\Delta)}{\hat{\pi}(x_i)} = \gamma_0(\theta,\Delta)+o_p(1),$$

$$(iii) \qquad \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i\dot{\psi}_{\theta}(x_i;\theta,\Delta)}{\hat{\pi}(x_i)} = \beta_0(\theta,\Delta)+o_p(1),$$

where $\sigma_0^2(\theta,\Delta) = \frac{1}{\mu}E_{F_1}\left(\frac{\psi^2(X;\theta,\Delta)}{\pi(X)}\right)$, $\beta_0(\theta,\Delta) = \frac{1}{\mu}E_{F_1}\left(\dot{\psi}_{\theta}(X;\theta,\Delta)\right)$ *and* $\gamma_0(\theta,\Delta) = \frac{1}{\mu}E_{F_1}\left(\dot{\psi}_{\Delta}(X;\theta,\Delta)\right)$.

Proof. Firstly,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 \le \frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\pi(x_i)}\right\}^2 + \frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 \sup_{0\le x_i\le x_{(n)}}\left|\frac{\pi(x_i)-\hat{\pi}(x_i)}{\pi(x_i)}\right|^2.$$

By the law of large numbers and Assumption (i), we have that

$$\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\pi(x_i)}\right\}^2 \quad\to\quad 2\sigma_0^2(\theta,\Delta)<\infty,$$

where $\sigma_0^2(\theta,\Delta) = E\left(\frac{\delta(X)\psi(X;\theta,\Delta)}{\pi(X)}\right)^2 = \frac{1}{\mu}E_{F_1}\left(\frac{\psi^2(X;\theta,\Delta)}{\pi(X)}\right)$.

According to [20],

$$\sup_{0\le x_i\le x_{(n)}}\left|\frac{\pi(x_i)-\hat{\pi}(x_i)}{\pi(x_i)}\right| = \sup_{0\le x_i\le x_{(n)}}\left|\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau_C}\frac{h(s,x_i)}{S_C(s)S_T(s)}dM_j^C(s)+o_p(n^{-1/2})\right| = o_P(1),$$

where $h(s,x_i)$ and $M_j^C(s)$ are defined in Theorem 3.1, therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 \le O_p(1)/(1-o_p(1)) = O_p(1).$$

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\pi(x_i)}\right\}^2 + \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\pi(x_i)-\hat{\pi}(x_i)}{\hat{\pi}(x_i)\pi(x_i)}\right\}^2 \delta_i^2\psi^2(x_i;\theta,\Delta) + \frac{2}{n}\sum_{i=1}^{n}\frac{\pi(x_i)-\hat{\pi}(x_i)}{\hat{\pi}(x_i)\pi^2(x_i)}\delta_i^2\psi^2(x_i;\theta,\Delta) = I_1+I_2+I_3,$$

and

$$\begin{aligned} I_1 &= \sigma_0^2(\theta,\Delta)+o_p(1),\\ I_2 &\le \sup_{0\le x_i\le x_{(n)}}\left|\frac{\pi(x_i)-\hat{\pi}(x_i)}{\pi(x_i)}\right|^2 \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2\\ &= o_p(1)O_p(1) = o_p(1),\\ I_3 &\le 2\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\pi(x_i)-\hat{\pi}(x_i)}{\pi(x_i)\hat{\pi}(x_i)}\psi(x_i;\theta\Delta)\delta_i\right)^2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\psi(x_i;\theta,\Delta)\delta_i\right)^2\right\}^{1/2} = 2(I_1I_2)^{1/2} = o_p(1). \end{aligned}$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right\}^2 = \sigma_0^2(\theta,\Delta)+o_p(1).$$

With the similar discussion, (ii) and (iii) can be also derived. □

Lemma A.3. *The following equality holds uniformly on $\{\theta: \ \|\theta-\theta_0\|\le cn^{-\varrho}\}$, where $1/3<\varrho<1/2$ and c is some constant:*

$$\lambda(\theta) = O_p(n^{-\varrho})$$

and

$$\lambda(\theta) = \frac{\sum_{i=1}^{n}w(x_i)\psi(x_i,\theta,\Delta)}{\sum_{i=1}^{n}\{w(x_i)\psi(x_i,\theta,\Delta)\}^2}+o_P(n^{-\varrho}),$$

where $\lambda(\theta)$ is determined by (7).

Proof. Denote $Z_i = \lambda(\theta)w(x_i)\psi(x_i,\theta,\Delta)$, and it is followed from (7) and the equality $\frac{1}{1+Z_i} = 1 - \frac{Z_i}{1+Z_i}$ that

$$\frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda(\theta)\{w(x_i)\psi(x_i,\theta,\Delta)\}^2}{1+\lambda(\theta)w(x_i)\psi(x_i,\theta,\Delta)},$$

thus

$$\frac{\lambda(\theta)}{n}\sum_{i=1}^{n}\{w(x_i)\psi(x_i,\theta,\Delta)\}^2 \le \frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta)(1+\lambda(\theta)\max_{1\le i\le n}(w(x_i)\psi(x_i,\theta,\Delta))),$$

which implies that

$$\lambda(\theta)\left\{\frac{1}{n}\sum_{i=1}^{n}\{w(x_i)\psi(x_i,\theta,\Delta)\}^2 - \frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta) \max_{1\le i\le n}\left|\frac{\delta_i\psi(x_i;\theta,\Delta)}{\hat{\pi}(x_i)}\right|\right\} \le \frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta). \tag{A.1}$$

By Assumption (ii) and a similar proof of Lemma 3 of [14], we obtain

$$\max_{1\le i\le n}\left|\frac{\delta_i\psi(x_i;\theta,\Delta)}{\pi(x_i)}\right| = o_p(n^{1/3}).$$

Therefore,

$$\max_{1\le i\le n}|w(x_i)\psi(x_i,\theta,\Delta)| \le \max_{1\le i\le n}\left|\frac{\delta_i\psi(x_i;\theta,\Delta)}{\pi(x_i)}\right| \left(1+\sup_{0\le x\le x_{(n)}}\left|\frac{\pi(x)-\hat{\pi}(x)}{\pi(x)}\right|\right) = o_p(n^{1/3}),$$

Moreover, by Lemma A.1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta_0,\Delta) \;\to\; N(0,\sigma_0^2(\theta_0,\Delta)),$$

and with Lemma A.2 (iii) and Assumption (ii), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta) &= \frac{1}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta_0,\Delta) + O_p(n^{-\varrho})\\ &= O_p(n^{-\varrho}).\end{aligned} \tag{A.2}$$

Combining (A.1), (A.2), (A.2) and Lemma A.2 (i), we have

$$\lambda(\theta) = O_P(n^{-\varrho}). \tag{A.3}$$

On the other hand, by applying Taylor expansion to (7), we can have

$$\lambda(\theta) = \frac{\sum_{i=1}^{n} w(x_i)\psi(x_i,\theta,\Delta)}{\sum_{i=1}^{n}(w(x_i)\psi(x_i,\theta,\Delta))^2} + R_n,$$

where

$$\begin{aligned}R_n &= \left\{\lambda^2(\theta,\Delta)\sum_{i=1}^{n}(w(x_i)\psi(x_i,\theta,\Delta))^3(1+o_P(1))\right\} \left\{\sum_{i=1}^{n}(w(x_i)\psi(x_i,\theta,\Delta))^2\right\}^{-1}\\ &\le \lambda^2(\theta,\Delta)\max_{1\le i\le n}|w(x_i)\psi(x_i,\theta,\Delta)|(1+o_P(1))\\ &= O_P(n^{-2\varrho})o_P(n^{1/3}) = o_P(n^{-\varrho}).\end{aligned}$$

□

Proof of Theorem 3.1. By Lemma A.3 and following the similar arguments to the proof of Lemma 1 of [16], the result can be easily proved. □

Proof of Theorem 3.2. By (10) and (11), the values of partial derivatives of $Q_{in}(\lambda,\theta)$ $(i=1,2,3)$ with respect to λ and θ at $(0,\theta_0)$ are as follows.

$$\begin{aligned}\frac{\partial Q_{1n}(0,\theta_0)}{\partial\lambda} &= -\frac{1}{n}\sum_{i=1}^{n}\left(w(x_i)\psi(x_i,\theta_0,\Delta)\right)^2;\\ \frac{\partial Q_{1n}(0,\theta_0)}{\partial\theta} &= \frac{1}{n}\sum_{i=1}^{n} w(x_i)\dot{\psi}_\theta(x_i,\theta_0,\Delta);\\ \frac{\partial Q_{2n}(0,\theta_0)}{\partial\lambda} &= \frac{1}{n}\sum_{i=1}^{n} w(x_i)\dot{\psi}_\theta(x_i,\theta_0,\Delta);\\ \frac{\partial Q_{2n}(0,\theta_0)}{\partial\theta^{\mathrm{T}}} &= -\frac{1}{n}\frac{\partial^2\ell_2(\theta_0)}{\partial\theta^{\mathrm{T}}\partial\theta}.\end{aligned}$$

The remaining proof can follow the similar discussion in Theorem 1 of [16]. □

Proof of Theorem 3.3. First, with the equality $Q_{2n}(0,\hat{\theta}_{\mathrm{MLE}},\Delta) = -\frac{\partial\ell_2(\hat{\theta}_{\mathrm{MLE}})}{\partial\theta} = 0$ and by applying Taylor expansion to (8), we have that

$$\begin{aligned}\frac{1}{n}\mathcal{L}(\hat{\theta}_{\mathrm{EL}},\Delta) &= \frac{\hat{\lambda}}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\hat{\theta}_{\mathrm{MLE}},\Delta)\\ &- \frac{\hat{\lambda}^2}{2n}\sum_{i=1}^{n}\Big(w(x_i)\psi(x_i,\hat{\theta}_{\mathrm{MLE}},\Delta)\Big)^2\\ &- \frac{1}{2n}(\hat{\theta}_{\mathrm{EL}}-\hat{\theta}_{\mathrm{MLE}})^{\mathrm{T}}\frac{\partial^2\ell_2(\hat{\theta}_{\mathrm{MLE}})}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_{\mathrm{EL}}-\hat{\theta}_{\mathrm{MLE}})\\ &+ o_P(1).\end{aligned}$$

(A.1) and (A.2) yield that

$$\begin{aligned}&\frac{\hat{\lambda}}{n}\sum_{i=1}^{n} w(x_i)\psi(x_i,\hat{\theta}_{\mathrm{MLE}},\Delta)\\ &= \frac{\hat{\lambda}^2}{n}\sum_{i=1}^{n}\Big(w(x_i)\psi(x_i,\hat{\theta}_{\mathrm{MLE}},\Delta)\Big)^2 + o_P(n^{-\varrho}).\end{aligned} \tag{A.4}$$

Once again, it follows from Taylor expansion that

$$\frac{1}{n}\frac{\partial\ell_2(\hat{\theta}_{\mathrm{EL}})}{\partial\theta} = \frac{1}{n}\frac{\partial\ell_2^2(\hat{\theta}_{\mathrm{MLE}})}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_{\mathrm{EL}}-\hat{\theta}_{\mathrm{MLE}}) + o_P(n^{-\varrho}).$$

On the other hand, it follows from $Q_{2n}(\hat{\lambda}, \hat{\theta}_{\mathrm{EL}}, \Delta) = 0$ that

$$\frac{1}{n}\frac{\partial \ell_2(\hat{\theta}_{\mathrm{EL}})}{\partial \theta} = \frac{\hat{\lambda}}{n}\sum_{i=1}^{n}\frac{w(x_i)\dot{\psi}_\theta(x_i, \hat{\theta}_{\mathrm{EL}}, \Delta)}{1+\hat{\lambda}w(x_i)\psi(x_i, \hat{\theta}_{\mathrm{EL}}, \Delta)} = \hat{\lambda}\beta_0 + o_p(1).$$

Therefore,

$$\text{(A.5)} \qquad \hat{\theta}_{\mathrm{EL}} - \hat{\theta}_{\mathrm{MLE}} = \hat{\lambda}\left\{\frac{1}{n}\frac{\partial^2 \ell_2(\hat{\theta}_{\mathrm{MLE}})}{\partial\theta\partial\theta^{\mathrm{T}}}\right\}^{-1}\beta_0 + o_p(1).$$

Combining (A.4), (A.5), Lemma A.2 and Theorem 3.2, we obtain that

$$\text{(A.6)} \qquad 2\mathcal{L}(\hat{\theta}_{\mathrm{EL}}, \Delta) = (\sqrt{n}\hat{\lambda})^2(\sigma_0^2 + \beta_0^{\mathrm{T}}\Sigma_0^{-1}\beta_0) + o_p(1).$$

Theorem 3.2 indicates that

$$2\mathcal{L}(\hat{\theta}_{\mathrm{EL}}, \Delta) \xrightarrow{\mathcal{D}} (C_\theta^2\Gamma + C_\theta^2\beta_0^{\mathrm{T}}\Sigma_0^{-1}\beta_0)(\sigma_0^2+\beta_0^{\mathrm{T}}\Sigma_0^{-1}\beta_0)\chi^2(1).$$

□

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, Dr. Heping Zhang, and the associate editor for their support during the review process. They would also like to thank two anonymous referees for their constructive suggestions and comments that greatly contributed to improving the final version of the manuscript. Li's work was supported by State Key Program of NNSF of China (91546202). Wu's work was supported by NNSF of China (61402276). Zhou's work was supported by the State Key Program of NNNSF of China (71331006) and the State Key Program in the Major Research Plan of NNSF of China (91546202).

Received 26 December 2016

REFERENCES

[1] Addona, V. and Wolfson, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow up. *Lifetime Data Analysis*, **12**, 267–284. MR2328577

[2] Asgharian, M. (2003). Biased sampling with right censoring: A note on Sun, Cui, Tiwari (2002). *The Canadian Journal of Statistics*, **31**, 349–350. MR2030129

[3] Asgharian, M., Wolfson, D. B., and Zhang, X. (2005). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine*, **25**, 1751–1767. MR2227351

[4] Bai, F., Huang, J., and Zhou, Y. (2014). Semiparametric estimation of treatment effects in two sample problems with censored data. *Statistica Sinica*, **24**, 121–146. MR3183677

[5] Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis*, **49**, 24–40. MR1275041

[6] DiCiccio, T. J., Hall, P. J. and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, **19**, 1053–1061. MR1105861

[7] Han, X., Small, S. D., Foster, P. D. and Patel, V. (2011). The effect of winning an Oscar Award on survival: Correcting for healthy performer survivor bias with a rank preserving structural accelerated failure time model. *The Annals of Applied Statistics*, **5**, 746–772. MR2840174

[8] He, S. and Wei, L. (2014). Empirical likelihood for right censored data with covariables. *Science China*, **57**, 1275–1286. MR3202016

[9] Hsieh, F., Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic cruve. *The Annals of Statistics*, **24**, 25–40. MR1389878

[10] Li, G., Tiwari, R. C., and Wells, M. T. (1996). Quantile comparison function in two-sample problems: with the application to comparisons of diagnostic markers. *Journal of the American Statistical Association*, **91**, 689–698. MR1395736

[11] Li, G., Tiwari, R. C., and Wells, M. T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika*, **86**, 487–502.

[12] Lin, C. and Zhou, Y. (2014). Inference for the treatment effects in two sample problems with right-censored and length-biased data. *Statistics & Probability Letters*, **90**, 17–24.

[13] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

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

[15] Qin, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Annals of the Institute of Statistical Mathematics*, **46**, 117–126.

[16] Qin, J. and Lawless, J. (1994). Empirical likelihood ratio and general estimating quations. *The Annals of Statistics*, **22**, 300–325.

[17] Qin, Y. S. (1997). Semi-parametric likelihood ratio confidence intervals for various differences of two populations. *Statistics & Probability Letters*, **33**, 135–143.

[18] Ren, J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*, **36**, 147–166.

[19] Redelmeier, D. A. and Singh, S. M. (2001). Survival in Academy Award-winning actors and actresses. *Annals of Internal Medicine*, **134**, 955–962.

[20] Shen, Y., Ning, J.and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association*, **104**, 1192–1202.

[21] Sylvestre, M. P., Huszti, E. and Hanley, J. A. (2006). Do Oscar winners live longer than less successful peers? A reanalysis of the evidence. *Annals of Internal Medicine*, **145**, 361–363.

[22] Thomas, D. and Grunkemeier, G. (1975). Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of the American Statistical Association*, **70**, 865–871.

[23] Zhao, Y. (2011). Empirical likelihood inference for the accelerated failure time model. *Statistics & Probability Letters*, **81**, 603–610.

[24] Zhou, Y. and Liang, H. (2005). Empirical likelihood-based semiparametric inference of the treatment effect in the two-sample problems with censoring. *Biometrika*, **92**, 271–282.

Tao Li
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai 200433
E-mail address: li.tao@mail.shufe.edu.cn

Mengyun Wu
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai 200433
E-mail address: wu.mengyun@mail.shufe.edu.cn

Yong Zhou
Academy of Mathematics and System Sciences
Chinese Academy of Science
Beijing 100190
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai 200433
E-mail address: yzhou@amss.ac.cn