Statistica Sinica **13**(2003), 51-68

EMPIRICAL LIKELIHOOD REGRESSION ANALYSIS FOR RIGHT CENSORED DATA

Gang Li[1] and Qi-Hua Wang[2]

[1] *University of California and*
[2] *Chinese Academy of Sciene and Heilongjiang University*

Abstract: Linear models are useful alternatives to the Cox (1972) proportional hazards model for analyzing censored regression data. This article develops empirical likelihood methods for linear regression analysis of right censored data. An adjusted empirical likelihood is constructed for the vector of regression coefficients using a synthetic data approach. The adjusted empirical likelihood is shown to have a central chi-squared limiting distribution, which enables one to make inference using standard chi-square tables. We also derive an adjusted empirical likelihood method for linear combinations of the regression coefficients. In addition, we discuss how to incorporate auxiliary information. A small simulation study is carried out to highlight the performance of the adjusted empirical likelihood methods compared with the traditional normal approximation method. It shows that the empirical likelihood confidence intervals tend to have more accurate coverage probabilities than the normal theroy intervals. An illustration is given using the Stanford Heart Transplant data.

Key words and phrases: Confidence intervals, linear models, right censoring, synthetic data.

1. Introduction

Owen (1991) and Chen (1993, 1994) derived empirical likelihood inference procedures for linear models. The empirical likelihood methods have sampling properties similar to those of the bootstrap and have wider validity than the usual parametric procedures. They also have the appealing feature that the shape and orientation of the resulting confidence regions are determined entirely by the data. In contrast, it would be difficult to determine the $1-\alpha$ central fraction of a point cloud with the bootstrap method when the dimension is two or higher.

In survival analysis, the survival time of interest is often not completely observed. For example, a common situation is that of right censoring where, due to the end of follow-up or occurrence of competing events, the survival times of some individuals are not observed, but are known to be greater than some observed values. The purpose of this paper is to extend the empirical likelihood method to linear regression analysis for right censored survival data. Linear models provide useful alternatives to the popular Cox (1972) model for analysis of survival data when the proportional hazards assumption fails to hold.

Specifically, assume that one observes n i.i.d. triples $(X_i, \tilde{Y}_i, \delta_i) = (X_i, Y_i \wedge C_i, I[Y_i \leq C_i])$, $i = 1, \ldots, n$, where for subject i, Y_i is a known monotone transformation of the survival time of interest, C_i is the corresponding censoring time and $X_i = (X_{i1}, \ldots, X_{ip})^\tau$ is a vector of p covariates. Consider the linear model

$$Y_i = X_i^\tau \beta_0 + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where β_0 is an (unknown) column vector of regression coefficients, $\epsilon_i = Y_i - E(Y_i|X_i)$, and C_i is independent of (X_i, Y_i), $i = 1, \ldots, n$. Note that the distribution of ϵ_i is completely unknown and that $\mathrm{Var}\,(\epsilon|X = x)$ is allowed to depend on x. In the absence of censoring, this model reduces to the linear model studied by Owen (1991) who gave a nice discussion of why empirical likelihood is adequate for linear models with heteroscedastic error terms (Owen (1991), Section 5.1). In this paper, we develop empirical likelihood inference for β_0 and its linear combinations based on right censored data $(X_i, \tilde{Y}_i, \delta_i)$, $i = 1, \ldots, n$.

Apparently the results of Owen (1991) and Chen (1993, 1994) for complete data do not apply to censored data since the Y_i's are not always observed. A possible solution is to consider a synthetic data approach, used by Koul, Susarla and Ryzin (1981), Leurgans (1987) Srinivasan and Zhou (1991) and Zhou (1992), to derive least squares estimates of β_0. For simplicity, we proceed with Koul et al.'s (1981) proposal. The basic idea is to first introduce a synthetic response variable whose expectation is close to that of Y_i. This is done in Section 2. A complete data empirical likelihood is then constructed for β_0 from the synthetic data as if they were i.i.d. observations. Because the synthetic data are in fact dependent, standard chi-square tables do not directly apply. By examining an asymptotic expansion of the empirical likelihood, we introduce an empirical adjustment and show that the adjusted empirical log-likelihood has an asymptotic standard chi-square distribution. The adjustment factor reflects the information loss due to censoring. In the absence of censoring, our adjusted empirical likelihood reduces to the standard one of Owen (1991). It is worth noting that a similar technique was used by Kitamura (1997) who developed a so-called block empirical likelihood using an adjustment factor in a dependent process model.

We also consider the problem of making empirical likelihood inference for linear combinations of the regression coefficients. Examples of linear combinations include a single coefficient, a subset of coefficients, and contrasts. For the complete data problem, Chen (1994) showed that the empirical likelihood still has a standard chi-square limiting distribution after the nuisance parameters are profiled out. However, it is not clear whether or not the same can be said for censored data. Moreover, profiling the nuisance parameters involves constrained optimization which may not be an easy task in high dimensional cases. Instead of using a profile likelihood, we replace the nuisance parameters by their least squares estimates and derive an adjusted empirical likelihood with an appropriate adjustment factor.

We further extend the adjusted empirical likelihood method to situations where there is available auxiliary information on X. The results are useful in instances where some population characteristics of the covariate X are known. For example, one may know the mean or median of X, or that the population distribution is symmetric about a known constant.

The use of a likelihood ratio in nonparametric settings dates back at least to Thomas and Grunkemeier (1975) who derived nonparametric likelihood ratio-based confidence intervals for survival probabilities. Its first theoretical development was due to Owen (1988, 1990), who introduced empirical likelihood confidence regions for the mean of a random vector based on i.i.d. observations. During the last decade, empirical likelihood has been extended to a wide range of applications including, among others, linear models (Owen (1991) and Chen (1993, 1994)), generalized linear models (Kolaczyk (1994), Chen and Cui (2002)), quantile estimation (Chen and Hall (1993), Zhou and Jing (2002)), biased sample models (Qin (1993)), generalized estimating equations (Qin and Lawless (1994)), truncation models (Li (1995a)), dependent process model (Kitamura (1997)), partial linear models (Wang and Jing (1999)), mixture proportions (Qin (1999)), random censorship models (Hollander, McKeague and Yang (1997), Li, Hollander, McKeague and Yang (1996), Adimari (1997), Li (1995b), Murphy (1995)), Li and Van Keilegom (2002), Wang and Li (2002), Wang and Jing (1999) and Wang and Wang (2001)), and confidence tubes for multiple quantile plots (Einmahl and McKeague (1999)). Some nice discussion of properties of empirical likelihood can be found in DiCiccio, Hall and Romano (1991), Hall (1992), and Hall and Scala (1990), and elsewhere.

The paper is organized as follows. In Section 2, we derive an adjusted empirical likelihood for making inference for β_0. A Wilks-type theorem is established. It ensures that the resulting adjusted empirical likelihood confidence region has asymptotically correct coverage probability. In Section 3, we develop an adjusted empirical likelihood method for combinations of the regression coefficients. In Section 4, we describe how to incorporate auxiliary information. In Section 5, we conduct a small simulation study to compare the adjusted empirical likelihood method with the traditional normal approximation method. An illustration is given using the Stanford Heart Transplant data. In Section 6, we note some limitations of the synthetic data approach and discuss possible extensions that could lead to better empirical likelihood procedures. Proofs are given in the appendix.

2. Adjusted Empirical Likelihood for Global Inference

In this section we derive an adjusted empirical likelihood (ADEL) method to make global inference for β_0. From now on, we assume that $E(X_i X_i^\tau)$ is positive definite.

Define a synthetic variable $Y_{iG} = \tilde{Y}_i \delta_i / (1 - G(\tilde{Y}_i -))$, $i = 1, \ldots, n$, where G is the cumulative distribution function of the censoring time C_i. It can be verified that $E(Y_{iG} \mid X_i) = E(Y_i \mid X_i)$. Hence, under the linear model (1), we have

$$Y_{iG} = X_i^\tau \beta_0 + e_i, \tag{2}$$

where $e_i = Y_{iG} - E(Y_{iG} \mid X_i)$.

It follows from (2) that $\beta_0 = (EX_i X_i^\tau)^{-1} E(X_i Y_{iG})$, or $EX_i(Y_{iG} - X_i^\tau \beta_0) = 0$. Therefore, for a given β, the problem of testing $H_0 : \beta_0 = \beta$ is equivalent to testing $E(W_i(\beta)) = 0$ based on n i.i.d. observations $W_i(\beta) = X_i(Y_{iG} - X_i^\tau \beta)$, $i = 1, \ldots, n$.

If G were known, one could test $EW_i(\beta) = 0$ using the empirical likelihood of Owen (1990):

$$l_n(\beta) = -2 \sup \left\{ \sum_{i=1}^n \log(np_i) \Big| \sum_{i=1}^n p_i W_i(\beta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0, i = 1, \ldots, n \right\}.$$

It follows from Owen (1990) that, under $H_0 : \beta_0 = \beta$, $l_n(\beta)$ has an asymptotic central chi-sqaure distribution with p degrees of freedom. An essential condition for this result to hold is that the $W_i(\beta)$'s in the linear constraint are i.i.d. random variables.

Unfortunately, the censoring distribution G is generally unknown and thus $l_n(\beta)$ cannot be computed since it depends on G. A natural solution is to replace G by its Kaplan-Meier (1958) estimator in $l_n(\beta)$. Specifically, let $W_{in}(\beta) = X_i(Y_{i\hat{G}n} - X_i^\tau \beta)$, where $\hat{G}_n(t)$ is the Kaplan-Meier estimator of G given by

$$1 - \hat{G}_n(t) = \prod_{i=1}^n \left[\frac{n-i}{n-i+1} \right]^{I[\tilde{Y}_{(i)} \leq t, \delta_{(i)} = 0]},$$

$\tilde{Y}_{(1)} \leq \cdots \leq \tilde{Y}_{(n)}$ are the order statistics of the $\tilde{Y}$-sample, and $\delta_{(i)}$ is the δ associated with $\tilde{Y}_{(i)}$, $i = 1, \ldots, n$. An estimated empirical log-likelihood is defined by

$$\tilde{l}_n(\beta) = -2 \sup \left\{ \sum_{i=1}^n \log(np_i) \Big| \sum_{i=1}^n p_i W_{in}(\beta) = 0, \sum_{i=1}^n p_i = 1, p_i \geq 0, i = 1, \ldots, n \right\}.$$

It is easy to show that

$$\tilde{l}_n(\beta) = 2 \sum_{i=1}^n \log\{1 + \lambda^\tau W_{in}(\beta)\}, \tag{3}$$

where λ is the solution of the equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_{in}(\beta)}{1+\lambda^{\tau}W_{in}(\beta)}=0. \tag{4}$$

Because $W_{in}(\beta)$, $i=1,\ldots,n$, are dependent, $\tilde{l}_n(\beta)$ no longer has an asymptotic standard chi-square distribution. In the appendix (Remark A.1), we show that $\tilde{l}_n(\beta_0)$ converges in distribution to $\sum_{i=1}^{p} w_l\chi_{1,l}^2$ where the $\chi_{1,l}^2$ are independent χ_1^2 random variables, the weights w_l are eigenvalues of $\Sigma_1^{-1}(\beta_0)(\Sigma_1(\beta_0)-\Sigma_2)$, and $\Sigma_1(\beta_0)$ and Σ_2 are defined by (C.Σ_1) and (C.Σ_2) in the appendix. Although the weights could be estimated from data, Monte Carlo simulations would be needed to compute percentiles of the limiting distribution even if the weights were known. Instead of estimating the distribution of $\tilde{l}_n(\beta)$ directly, we present another method by introducing an adjustment factor for $\tilde{l}_n(\beta)$ so that the adjusted empirical likelihood function has an asymptotic standard chi-square distribution.

The following notations are needed. Let F denote the distribution of Y_i. Let $\hat{F}_n$ be the Kaplan-Meier estimator of F. Let $Q_n(s)=(\sum_{i=1}^{n} I[\tilde{Y}_i\le s])/n$,

$$\begin{aligned}
H_n(s) &= \frac{\frac{1}{n}\sum_{i=1}^{n} X_iY_{i\hat{G}_n}I[s<\tilde{Y}_i]}{(1-\hat{G}_n(s))(1-\hat{F}_n(s-))},\\
\Lambda_n^{\hat{G}_n}(t) &= \int_0^t \frac{1}{1-\hat{G}_n(s-)}d\hat{G}_n(s)=\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_i)I[\tilde{Y}_i\le t]}{1-Q_n(\tilde{Y}_i)},\\
\hat{\Sigma}_{1n}(\beta) &= \frac{1}{n}\sum_{i=1}^{n} X_iX_i^{\tau}(Y_{i\hat{G}_n}-X_i^{\tau}\beta)^2,\\
\hat{\Sigma}_{2n} &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(H_n(\tilde{Y}_i)H_n^{\tau}(\tilde{Y}_i)(1-\triangle\Lambda_n^{\hat{G}_n}(\tilde{Y}_i)),\\
\hat{\Sigma}_n(\beta) &= \hat{\Sigma}_{1n}(\beta)-\hat{\Sigma}_{2n},\\
S_n(\beta) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{in}(\beta)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{in}(\beta)\right)^{\tau}.
\end{aligned}$$

The adjusted empirical likelihood function is defined by

$$\hat{l}_{n,ad}(\beta)=r_n(\beta)\tilde{l}_n(\beta), \tag{5}$$

where $r_n(\beta)=\operatorname{tr}(\hat{\Sigma}_n^{-1}(\beta)S_n(\beta))/\operatorname{tr}(\hat{\Sigma}_{1n}^{-1}(\beta)S_n(\beta))$. The adjustment factor can be derived by examining the leading term in the asymptotic expansion of $\tilde{l}_n(\beta)$ given in the proof of Theorem 2.1 in the appendix. Another way of motivating $r_n(\beta)$ is to use a result of Rao and Scott (1981) who showed that the distribution of $\tilde{r}(\beta_0)\sum_{i=1}^{p} w_i\chi_{1,i}^2$ may be approximated by χ_p^2, where $\tilde{r}(\beta_0)=$

$p/\operatorname{tr}\{\Sigma_1^{-1}(\beta_0)\Sigma(\beta_0)\}$, $\operatorname{tr}(\cdot)$ denoting the trace operator. Note that $\tilde r(\beta_0)$ can also be written as $\tilde r(\beta_0) = \operatorname{tr}\{\Sigma^{-1}(\beta_0)\Sigma(\beta_0)\}/\operatorname{tr}\{\Sigma_1^{-1}(\beta_0)\Sigma(\beta_0)\}$. Replacing $\Sigma^{-1}(\beta_0), \Sigma_1^{-1}(\beta_0)$ and $\Sigma(\beta_0)$ in $\tilde r(\beta_0)$ by the sample estimates $\hat\Sigma_n^{-1}(\beta_0)$, $\hat\Sigma_{1n}^{-1}(\beta_0)$ and $S_n(\beta_0)$, respectively, leads to the expression for $r_n(\beta_0)$.

Theorem 2.1. *Assume that the conditions* (C.XY), (C.FG), (C.Σ_1) *and* (C.Σ_2) *listed in the appendix hold.*

(a) *As* $n \to \infty$, $\hat l_{n,ad}(\beta_0) \xrightarrow{d} \chi_p^2$, *where* χ_p^2 *is a standard chi-square random variable with* p *degrees of freedom.*

(b) *For* $0 < \alpha < 1$, *define* $I_\alpha = \{\beta : \hat l_{n,ad}(\beta) \le \chi^2_{p,\alpha}\}$, *where* $\chi^2_{p,\alpha}$ *is the upper* α*th percentile of the* χ_p^2 *distribution. Then* $\lim_{n\to\infty} P(\beta_0 \in I_a) = 1-\alpha$.

(c) *Define* $I_e = \{\beta : \tilde l_n(\beta) \le \chi^2_{p,\alpha}\}$. *Then* $\lim_{n\to\infty} P(\beta_0 \in I_e) > 1-\alpha$.

Remark 2.1. It can be shown that, under mild conditions, $r_n(\beta)$ is always greater than or equal to 1. This implies that the adjusted interval I_a is a subset of the unadjusted interval I_e. Furthermore, if $\delta_i = 1$ for all i, then $r_n(\beta) \equiv 1$ and $Y_{i\hat G_n} = Y_i$ for all i. Therefore the adjusted empirical likelihood reduces to Owen's (1991) empirical likelihood in the absence of censoring.

Remark 2.2. Theorem 2.1 (b)-(c) show that the asymptotic confidence level of I_a is $1-\alpha$ and that of I_e exceeds $1-\alpha$. In Section 5 we present a simulation study which indicates that, for small samples, the actual confidence level of I_a tends to be lower than $1-\alpha$, especially when there is heavy censoring. In such cases, I_e showed smaller coverage probability errors than I_a since it is wider than I_a. For large samples, however, I_a is preferred to I_e since it is narrower and is expected to have a smaller coverage error, Theorem 2.1 (b)-(c) and simulation.

3. Adjusted Empirical Likelihood for Linear Combinations of β_0

This section extends the adjusted empirical likelihood method to make inference for a vector of linear combinations $\theta_0 = C\beta_0$ of β_0, where $C = (C_1, C_2)$, C_1 is a $k \times k$ matrix and C_2 is a $k \times (p-k)$ matrix ($k \le p-1$). For example, θ_0 is the subvector of the first k regression coefficients if $C_1 = I_k$ and $C_2 = 0$. If $k = 1$, then θ_0 reduces to a single linear combination, which includes an individual regression coefficient and the mean response at a given X level as special cases. Without loss of generality, we assume C_1^{-1} exists.

Let $\gamma_0 = (\theta_0^\tau, \beta_{0(k)}^\tau)^\tau$, where $\beta_{0(k)}$ denotes the column subvector of the last $p-k$ elements of β_0. Write $X_i = (\boldsymbol{X}_{i1}^\tau, \boldsymbol{X}_{i2}^\tau)^\tau$, where $\boldsymbol{X}_{i1}$ and $\boldsymbol{X}_{i2}$ are $k\times 1$ and $(p-k)\times 1$ subvectors. Let $\tilde{\boldsymbol{X}}_i = (\tilde{\boldsymbol{X}}_{i1}^\tau, \tilde{\boldsymbol{X}}_{i2}^\tau) = (\tilde{\boldsymbol{X}}_{i1}^\tau C_1^{-1}, \tilde{\boldsymbol{X}}_{i2}^\tau - \tilde{\boldsymbol{X}}_{i1}^\tau C_1^{-1} C_2)^\tau$. Then, model (1) reduces to $Y_i = \tilde{\boldsymbol{X}}_i^\tau \gamma_0 + \epsilon_i$, $i = 1, \ldots, n$.

Let $\hat\gamma_n(G) = (\sum_{i=1}^n \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^\tau)^{-1}(\sum_{i=1}^n \tilde{\boldsymbol{X}}_i Y_{iG})$ and let $\hat\beta_{n(k)}(G)$ denote the subvector of the last $p-k$ elements of $\hat\gamma_n(G)$. Note that $E\{\tilde{\boldsymbol{X}}_{i1}(Y_{iG} - \tilde{\boldsymbol{X}}_{i1}^\tau\theta_0 -$

$\tilde{\boldsymbol{X}}_{i2}^{\tau}\hat{\beta}_{n(k)}(G))\} = 0$, $i = 1, \ldots, n$. Similar to the previous section, for a given θ, we introduce the auxiliary variables $u_{in}(\theta) = \tilde{\boldsymbol{X}}_{i1}(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_{il}^{\tau}\theta - \tilde{\boldsymbol{X}}_{i2}^{\tau}\hat{\beta}_{n(k)}(\hat{G}_n))$, $i = 1, \ldots, n$, and define an estimated empirical likelihood function $l_{nk}(\theta) = 2\sum_{i=1}^{n}\log(1 + \lambda^{\tau}u_{in}(\theta))$, where λ satisfies $\sum_{i=1}^{n}u_{in}(\theta)/[1 + \lambda^{\tau}u_{in}(\theta)] = 0$.

Again, an adjustment factor is needed for $l_{nk}(\theta_0)$ to have a central chi-square limiting distribution. Write

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^{\tau} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_{i1}\tilde{\boldsymbol{X}}_{i1}^{\tau}, & K_n^{\tau} \\ K_n, & P_n \end{pmatrix}.$$

Let

$$\begin{aligned}
\eta_{ni} &= \tilde{\boldsymbol{X}}_{i1} - \left(\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j1}\tilde{\boldsymbol{X}}_j^{\tau}\right)\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_j\tilde{\boldsymbol{X}}_j^{\tau}\right)^{-1}\tilde{\boldsymbol{X}}_i \\
&\quad + \left(\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j1}\tilde{\boldsymbol{X}}_{j1}^{\tau}\right)\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j1}\tilde{\boldsymbol{X}}_{j1}^{\tau} - K_n^{\tau}P_n^{-1}K_n\right)^{-1}(\tilde{\boldsymbol{X}}_{i1} - K_n^{\tau}P_n^{-1}\tilde{\boldsymbol{X}}_{i2}), \\
H_{n0}(s) &= \frac{\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}Y_{i\hat{G}_n}I[s < \tilde{Y}_i]}{(1 - \hat{G}_n(s))(1 - \hat{F}_n(s-))}, \\
\hat{\Sigma}_{10n}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\eta_{ni}\eta_{ni}^{\tau}(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_{i1}^{\tau}\theta - \tilde{\boldsymbol{X}}_{i2}^{\tau}\hat{\beta}_{n(k)}(\hat{G}_n))^2, \\
\hat{\Sigma}_{20n} &= \frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i)H_{n0}(\tilde{Y}_i)H_{n0}^{\tau}(\tilde{Y}_i)(1 - \triangle\Lambda_n^{\hat{G}_n}(\tilde{Y}_i)), \\
\hat{\Sigma}_{n0}(\theta) &= \hat{\Sigma}_{10n}(\theta) - \hat{\Sigma}_{20n}, \\
\tilde{\Sigma}_{n0}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}u_{in}(\theta)u_{in}(\theta)^{\tau}, \\
S_{n0}(\theta) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_{in}(\theta)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_{in}(\theta)\right)^{\tau}.
\end{aligned}$$

An adjusted empirical likelihood is then defined by $l_{nk,ad}(\theta) = r_{n0}l_{nk}(\theta)$, where $r_{n0}(\theta) = \mathrm{tr}(\hat{\Sigma}_{n0}^{-1}(\theta)S_{n0}(\theta))/\,\mathrm{tr}(\tilde{\Sigma}_{n0}^{-1}(\theta)S_{n0}(\theta))$.

Theorem 3.1. *Assume that* (C.XY) *and* (C.FG) *given in the appendix hold. In addition, assume that* $\Sigma_{10}(\theta_0) = E[\eta_1\eta_1^{\tau}(Y_{1G} - \tilde{X}_1^{\tau}\gamma_0)^2]$ *and* $\Sigma_{20} = \int_0^{\tau_Q}H_0(s)H_0^{\tau}(s)\times(1 - F(s-))(1 - \triangle\Lambda^G(s))dG(s)$ *are positive definite matrices, where*

$$\begin{aligned}
H_0(s) &= \frac{E[\eta_1 Y_{1G}I(s < \tilde{Y}_1)]}{(1 - G(s))(1 - F(s-))}, \\
\eta_i &= \tilde{\boldsymbol{X}}_{i1} - E(\tilde{\boldsymbol{X}}_{11}\tilde{\boldsymbol{X}}_1^{\tau})\{E(\tilde{\boldsymbol{X}}_1\tilde{\boldsymbol{X}}_1^{\tau})\}^{-1}\tilde{\boldsymbol{X}}_i
\end{aligned}$$

$$+E(\tilde{\boldsymbol{X}}_{11}\tilde{\boldsymbol{X}}_{11}^{\tau})\{E(\tilde{\boldsymbol{X}}_{11}\tilde{\boldsymbol{X}}_{11}^{\tau})-K^{\tau}P^{-1}K\}^{-1}(\tilde{\boldsymbol{X}}_{i1}-K^{\tau}P^{-1}\tilde{\boldsymbol{X}}_{i2}),$$
$$K=E(K_n),\qquad P=E(P_n).$$

Then, $l_{nk,ad}(\theta_0)$ *is asymptotically standard chi-square distributed with* k *degrees of freedom.*

4. Adjusted Empirical Likelihood with Auxiliary Information

In many applications, some auxiliary population characteristics of X are known. It has been shown in the literature that effective usage of the available auxiliary information can lead to more efficient inferences; see e.g., Chen and Qin (1993), Qin and Lawless (1994) and Zhang (1995, 1996). In this section we show how to incorporate auxiliary information of X using an adjusted empirical likelihood.

Assume that the available auxiliary information on X is given in the form $Eg(X)=0$, where $g(x)=(g_1(x),\ldots,g_r(x))^{\tau}$, $r\geq 1$, is a vector of r known functions.

To make use of the auxiliary information, we maximize

$$\prod_{i=1}^{n} p_i \tag{6}$$

subject to $\sum_{i=1}^n p_i=1$, $\sum_{i=1}^n p_i g(X_i)=0$ and $\sum_{i=1}^n p_i\psi_{in}(\xi)=0$, where $\psi_{in}(\xi)$ is $W_{in}(\beta)$ or $u_{in}(\theta)$ as in Sections 2 and 3, depending on the context.

Let $A_{ni}(\xi)=(g^{\tau}(X_i),\psi_{in}^{\tau}(\xi))^{\tau}$. By the method of Lagrange multipliers, (6) is maximized at $p_{in}=1/[n(1+\zeta_n^{\tau}A_{ni}(\xi))]$, $i=1,\ldots,n$, where ζ_n satisfies $\sum_{i=1}^n A_{ni}(\xi)/[n(1+\zeta_n^{\tau}A_{ni}(\xi))]=0$. Hence, the empirical log-likelihood ratio function is given by $l_{n,AU}(\xi)=-2\sum_{i=1}^n\log np_{in}=2\sum_{i=1}^n\log(1+\zeta_n^{\tau}A_{ni}(\xi))$.

Just like $\tilde{l}_n(\beta)$ in Section 2, an adjustment factor is needed for $l_{n,AU}(\xi)$ to have a standard chi-square asymptotic distribution. Let $V_{n1}(\xi)=(\sum_{i=1}^n g(X_i)g^{\tau}(X_i))/n$, $V_{n2}(\xi)=(\sum_{i=1}^n g(X_i)\psi_{in}(\xi))/n$ and $V_{n3}(\xi)=(\sum_{i=1}^n\psi_{in}(\xi)\psi_{in}^{\tau}(\xi))/n$,

$$V_{n1,AU}(\beta)=\begin{pmatrix}V_{n1}(\xi), & V_{n2}(\xi)\\ V_{n2}^{\tau}(\xi), & V_{n3}(\xi)\end{pmatrix}\quad\text{and}\quad V_{n2,AU}(\beta)=\begin{pmatrix}V_{n1}(\xi), & 0\\ 0 & \hat{S}_n(\xi)\end{pmatrix},$$

where $\hat{S}_n(\xi)$ is $\hat{\Sigma}_n(\beta)$ defined in Section 2 or $\hat{\sigma}_n^2(\theta)$ in Section 3.

Similar to (5), we define an adjusted empirical log-likelihood for ξ by $\hat{l}_{n,AU}(\xi)=r_{n,AU}(\xi)l_{n,AU}(\xi)$, where $r_{n,AU}(\xi)=\operatorname{tr}(V_{n2,AU}^{-1}(\xi)\Psi_n(\xi))/\operatorname{tr}(V_{n1,AU}^{-1}(\xi)\Psi_n(\xi))$ and $\Psi_n(\xi)=(\sum_{i=1}^n A_{ni}(\xi))(\sum_{i=1}^n A_{ni}(\xi))^{\tau}$.

Theorem 4.1. *Assume that* $Eg(X)g^{\tau}(X)$ *is positive definite and that* $E\frac{g(X)X^{\tau}\tilde{Y}\delta}{1-G(\tilde{Y})}$ *exists.*

(a) *Let* $\xi = \beta$ *and* $\psi_{in}(\xi) = W_{in}(\beta)$. *Then, under the conditions of Theorem* 2.1, $\hat{l}_{n,AU}(\beta_0) \xrightarrow{d} \chi^2_{p+r}$ *as* $n \to \infty$.

(b) *Let* $\psi_{in}(\xi) = u_{in}(\theta)$, *where* $\xi = \theta$ *is a vector of* k *linear combinations of* β *as in Section* 3. *Then, under conditions of Theorem* 3.1, $\hat{l}_{n,AU}(\theta_0) \xrightarrow{d} \chi^2_{r+k}$ *as* $n \to \infty$.

5. Example and Simulation

In this section we illustrate the adjusted empirical likelihood method and compare it to the normal approximation method using a real data set. We also present a small simulation study to compare the performance of empirical likelihood confidence intervals to the normal approximation method.

Consider the heart transplant data of Miller ((1976), Table 1). The data includes the lengths of survival (in days) after transplantation, ages at time of transplant, and T5 mismatch scores for 69 patients who received heart transplants at Stanford and were followed from October 1 1967 to April 1 1974. The T5 mismatch score is a measure of the degree of dissimilarity between the donor and recipient tissue. Twenty-four patients were still alive on April 1 1974 and thus their survival times were censored.

Let Y be the logarithm to base 10 of the length of survival from transplantation. The three models we consider are (I) regress Y on the mismatch score T5; (II) regress Y on age; (III) regress Y on both T5 and age (Koul, Susarla and van Ryzin (1981)). As in Koul et al. (1981), regressions of survival on the mismatch score T5 were performed with nonrejection related death being treated as censoring since the mismatch score is directed at the rejection phenomenon (Miller (1976)). Confidence intervals for the slope parameters based on the normal approximation method (cf. Koul et al. (1981) and Lai, Ying and Zheng (1995)) the adjusted empirical likelihood (ADEL) and the estimated likelihood are given in Table 1.

Table 1. 95% adjusted empirical likelihood (ADEL), estimated empirical likelihood (EEL), and normal confidence interval estimates for heart transplant data.

			Confidence Intervals		
Model	Parameter	Estimate	ADEL	EEL	Normal
(I)	β_{T5}	0.258	$[-0.512, 0.943]$	$[-0.518, 0.947]$	$[-0.466, 0.943]$
(II)	β_{age}	0.054	$[0.019, 0.108]$	$[0.016, 0.112]$	$[0.017, 0.096]$
(III)	β_{T5}	0.052	$[-0.717, 0.759]$	$[-0.721, 0.762]$	$[-0.643, 0.746]$
	β_{age}	0.077	$[0.042, 0.139]$	$[0.037, 0.150]$	$[0.039, 0.114]$
(IV)	β_{T5}	-0.108	$[-0.691, 0.405]$	$[-0.696, 0.409]$	$[-0.601, 0.386]$
	β_{age}	0.056	$[0.021, 0.109]$	$[0.018, 0.113]$	$[0.019, 0.093]$

It is seen that for the heart transplant data, the empirical confidence intervals are comparable to those of the normal approximation method. As expected, the estimated empirical likelihood (EEL) is the most conservative and gives larger intervals. It is observed that the empirical likelihood confidence intervals are asymmetric about the point estimate. They are shifted to the left for β_{T5} and to the right for β_{age} compared to the normal confidence intervals. Recall that the traditional normal approximation method always imposes symmetry on the confidence interval. This is not a desirable property since the underlying distribution of the parameter estimate can be skewed. The empirical likelihood method is able to pick up possible skewness in the underlying distribution of the parameter estimate.

We carried out a small Monte Carlo simulation to examine coverage probabilities of the empirical likelihood confidence intervals compared to the normal approximation method. The data were generated from the following model: $Y = 1 + X + \epsilon$, where X and ϵ are independent normal random variables with mean 0 and variance 0.25, the censoring time C is a normal random variable with mean μ and standard deviation 4. We vary μ to produce different amounts of censoring. We also vary the sample size n. The simulated confidence levels of the empirical likelihood and normal confidence intervals for the slope parameter are given in Table 2. Each entry in the table was computed using 5000 Monte Carlo samples.

Table 2. Simulated coverage probabilities (%) of the normal, the adjusted empirical likelihood (ADEL) and the estimated empirical likelihood (EEL) confidence intervals for the slope parameter (nominal level = $1 - \alpha$).

Censoring	Sample	$1-\alpha = 90\%$			$1-\alpha = 95\%$		
Rate	Size	Normal	ADEL	EEL	Normal	ADEL	EEL
60%	50	77.9*	83.5	86.6	82.6*	88.8	92.0
	100	82.2*	86.5	90.5	88.5*	92.4	94.7
	500	90.4	90.9	94.4*	94.2	95.1	97.2*
32%	50	85.6*	87.8	89.5	87.4*	92.1	93.7
	100	89.4*	89.9	91.6	93.0*	94.8	96.4
	500	91.3	91.4	94.1*	95.4	95.6	97.1*

(* indicates a coverage probability that deviates the most from the nominal level among the three methods: "normal", "ADEL" and "EEL".)

Table 2 shows that the adjusted empirical likelihood confidence interval has more accurate coverage probabilities than the normal approximation method. The improvement of the empirical likelihood method is usually more pronounced for small samples (e.g., $n = 50$). Although the estimated empirical likelihood method is conservative for large samples (e.g., $n = 500$), it does improve the coverage for small samples. We actually conducted a more extensive simulation. The results are consistent with those in Table 2 and thus are not included.

6. Concluding Remarks

Empirical likelihood methods are studied for linear regression analysis of right censored data. Our results are valid for the heteroscedastic model that allows the conditional variance of the response to vary with the level of the covariate. The proposed method for the coefficient vector β_0 reduces to that of Owen (1991) in the absence of censoring. However, our method for linear combinations of β_0 has not been seen in the literature even for complete data. The empirical likelihood method demonstrates better small sample performance than the normal approximation method in a small simulation study and the improvement could be more pronounced in higher dimensional cases.

The synthetic data used in this article has some limitations. As shown by our simulation results (Table 2), the coverage probability can be far below the nominal level when there is heavy censoring and the sample size is small, even though the adjusted empirical likelihood method produces some improvement over the normal approximation method. Other types of synthetic data have been suggested in the literature for right censored data; see, e.g., Leurgans (1987) and Lai et al. (1995) among others. We point out that the idea used in this article can be applied directly to derive adjusted empirical likelihood procedures based on other synthetic data. Furthermore, one could develop adjusted empirical likelihood along the same lines based on general estimating equations (Buckley and James (1979) and Lai et al. (1995)). The use of other types of synthetic data or general estimating equations may lead to more efficient empirical likelihood procedures. Further investigation of various adjusted empirical likelihood methods for right censored data will be carried out in another paper.

7. Appendix

The following conditions are needed in Theorem 2.1.

(C.XY). $E(XI[s<Y])$ exists for every $0 \le s < \infty$.

(C.FG) (i). For all $s \le \tau_Q \equiv \inf\{t : Q(t)=1\}$, $G(s)$ and $F(s)$ have no common jumps, where $Q(t) = P(\tilde{Y} \le t)$.

(ii). $E\Big[\frac{\|X\|Y}{(1-G(y))(1-F(Y))^{\frac{1}{2}}}\Big] < \infty$.

(iii). $\int_0^{\tau_Q} \|H(s)\|\{[1-F(s)]/[1-F(s-)]\}[dG(s)/(1-G(s))] < \infty$, where $H(s) = E[XY_G I(s<\tilde{Y})]/\{(1-G(s))(1-F(s-))\}$.

(C.Σ_1). $\Sigma_1(\beta_0) = E[XX^\tau(Y_G - X^\tau\beta_0)^2]$ is a positive definite matrix.

(C.Σ_2). $\Sigma_2 = \int_0^\infty H(s)\,H^\tau(s)\,(1-F(s-))(1-\triangle\Lambda^G(s))\,dG(s)$ is a positive definite matrix, where

$$\Lambda^G(t) = \int_{-\infty}^{t} \frac{1}{1-G(s-)} dG(s).$$

The following lemma is needed to prove Theorem 2.1. For simplicity, we denote $W_{in}(\beta_0)$ by W_{in}.

Lemma A.1. *Suppose that the assumptions* (C.XY), (C.FG)(ii), (iii) *and* (C.Σ_1) *hold. Then* $n^{-\frac{1}{2}}\sum_{i=1}^{n} W_{in} \xrightarrow{\mathcal{L}} Z$, *where* Z *is a p-variate normal* $N(0, \Sigma(\beta_0))$ *random vector,* $\Sigma(\beta_0) = \Sigma_1(\beta_0) - \Sigma_2$, *and* $\Sigma_1(\beta_0)$ *and* Σ_2 *are defined in assumptions* (C.Σ_1) *and* (C.Σ_2).

Proof. The proof is similar to that of Theorem 2 of Lai, Ying and Zheng (1995), and is omitted.

Proof of Theorem 2.1. (a) To prove part (a) of the theorem, we need to show that (i) $\max_{1\le i\le n}\|W_{in}\| = o_p(n^{\frac{1}{2}})$, and (ii) $\lambda = O_p(n^{-\frac{1}{2}})$. We first prove (i). It is seen that

$$\max_{1\le i\le n}\|W_{in}\| \le \max_{1\le i\le n}\|X_i(Y_{i\hat{G}_n} - Y_{iG})\| + \max_{1\le i\le n}\|W_i\|, \tag{7}$$

$$\max_{1\le i\le n}\|X_i(Y_{i\hat{G}_n} - Y_{iG})\| \le \max_{1\le i\le n}\|X_iY_{iG}\| \sup_{0\le z\le \tilde{Y}_{(n)}}\left|\frac{\hat{G}_n(z)-G(z)}{1-\hat{G}_n(z)}\right|. \tag{8}$$

By Lemma 3 of Owen (1990), we have

$$\max_{1\le i\le n}\|W_i\| = o(n^{\frac{1}{2}}) \tag{9}$$

under assumption (C.FG)(ii). Moreover, it follows from Zhou (1992) that

$$\sup_{0\le s\le \tilde{Y}_{(n)}}\left|\frac{\hat{G}_n(s-)-G(s-)}{1-\hat{G}_n(s-)}\right| = O_p(1). \tag{10}$$

Thus, (i) follows immediately from (7)−(10).

Next, we prove (ii). Let $\lambda = \rho\theta$, where $\rho \ge 0$ and $\|\theta\| = 1$. Recall that $\hat{\Sigma}_{1n}(\beta_0) = 1/n\sum_{i=1}^{n} W_{in}W_{in}^{\tau}$. Let

$$\tilde{\Sigma}_{1n}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} X_iX_i^{\tau}\left(\frac{\delta_i\tilde{Y}_i}{1-G(\tilde{Y}_i)} - X_i^{\tau}\beta_0\right)^2.$$

It can be shown that

$$\hat{\Sigma}_{1n}(\beta_0) = \tilde{\Sigma}_{1n}(\beta_0) + o_p(1). \tag{11}$$

Let σ_p be the smallest eigenvalue of $S = E[XX^{\tau}(\tilde{Y}_G - X^{\tau}\beta_0)^2]$. Then, by Owen (1990),

$$\theta'\tilde{\Sigma}_{1n}(\beta_0)\theta \ge \sigma_p + o_p(1). \tag{12}$$

Let e_j be the unit vector in the jth coordinate direction. By Lemma A.1,

$$\left\| \frac{1}{n} \sum_{j=1}^{p} e_j' \sum_{i=1}^{n} W_{in} \right\| = O_p(n^{-\frac{1}{2}}). \tag{13}$$

It follows from (4), (11)−(13), and the arguments used in the proof of (2.14) of Owen (1990) that $\|\lambda\| = O_p(n^{-\frac{1}{2}})$. This proves (ii).

It follows from (i) and (ii) that $\max_{1\le i\le n} |\lambda^\tau W_{in}| = O_p(n^{-\frac{1}{2}}) o_p(n^{\frac{1}{2}}) = o_p(1)$. Hence, by Taylor's expansion, we have $\log(1+\lambda^\tau W_{in}) = \lambda^\tau W_{in} - \frac{1}{2}(\lambda^\tau W_{in})^2 + \eta_i$, where, for some constant $C > 0$, $P(|\eta_i| \le C|\lambda^\tau W_{in}|^3, 1 \le i \le n) \to 1$ as $n \to \infty$. Therefore

$$\tilde{l}_n(\beta_0) = 2\sum_{i=1}^{n} \log\{1 + \lambda^\tau W_{in}\} = 2\sum_{i=1}^{n} \left(\lambda^\tau W_{in} - \frac{1}{2}(\lambda^\tau W_{in})^2\right) + r_n, \tag{14}$$

$$P\left(|r_n| \le C\sum_{i=1}^{n} |\lambda^\tau W_{in}|^3\right) \to 1, \quad \text{as } n \to \infty. \tag{15}$$

Note that $\sum_{i=1}^{n} |\lambda^\tau W_{in}|^3 \le \|\lambda\|^3 \max_{1\le i\le n} \|W_{in}\| \sum_{i=1}^{n} \|W_{in}\|^2 = o_p(1)$, where the last step follows from (i), (ii), and the fact that

$$\frac{1}{n}\sum_{i=1}^{n} \|W_{in}\|^2 = O_p(1). \tag{16}$$

This, combined with (15), implies that

$$|r_n| = o_p(1). \tag{17}$$

Note that

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} \frac{W_{in}}{1+\lambda W_{in}} \\ &= \frac{1}{n}\sum_{i=1}^{n} W_{in}\left[1 - \lambda^\tau W_{in} + \frac{(\lambda^\tau W_{in})^2}{1+\lambda^\tau W_{in}}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} W_{in} - \left(\frac{1}{n}\sum_{i=1}^{n} W_{in}W_{in}^\tau\right)\lambda + \frac{1}{n}\sum_{i=1}^{n} \frac{W_{in}(\lambda^\tau W_{in})^2}{1+\lambda^\tau W_{in}}. \end{aligned} \tag{18}$$

By (11), (18), (i) and (ii), we get

$$\lambda = \left(\sum_{i=1}^{n} W_{in}W_{in}^\tau\right)^{-1} \sum_{i=1}^{n} W_{in} + o_p(n^{-\frac{1}{2}}). \tag{19}$$

It is seen from (18) that

$$0=\sum_{i=1}^{n}\frac{\lambda^\tau W_{in}}{1+\lambda^\tau W_{in}}=\sum_{i=1}^{n}(\lambda^\tau W_{in})-\sum_{i=1}^{n}(\lambda^\tau W_{in})^2+\frac{1}{n}\sum_{i=1}^{n}\frac{(\lambda^\tau W_{in})^3}{1+\lambda^\tau W_{in}}. \tag{20}$$

Moreover, by (i), (ii) and (16), we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(\lambda^\tau W_{in})^3}{1+\lambda^\tau W_{in}}=o_p(1). \tag{21}$$

It is easy to see that (20) and (21) imply

$$\sum_{i=1}^{n}\lambda^\tau W_{in}=\sum_{i=1}^{n}(\lambda^\tau W_{in})^2+o_p(1). \tag{22}$$

Combining (14), (17), (19) and (22) yields

$$\tilde{l}_n(\beta_0)=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)^\tau\hat{\Sigma}_{1n}^{-1}(\beta_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)+o_p(1). \tag{23}$$

A direct argument can be used to prove

$$\hat{\Sigma}_{1n}(\beta_0)\xrightarrow{p}\Sigma_1(\beta_0), \tag{24}$$

where $\Sigma_1(\beta_0)$ is defined in assumption (C.Σ_1) in the beginning of this section. Furthermore,

$$\hat{\Sigma}_{2n}(\beta_0)=\int_0^\infty H_n(z)H_n^\tau(z)(1-\triangle\Lambda_n^{\hat{G}_n}(z))(1-Q_n(z-))d\Lambda^{\hat{G}_n}(z)\xrightarrow{p}\Sigma_2(\beta_0),$$

by Stute and Wang (1993). Hence,

$$\hat{\Sigma}_n(\beta_0)\xrightarrow{p}\Sigma(\beta_0). \tag{25}$$

By Lemma A.1, (17), (23) and (25),

$$\hat{l}_{n,ad}(\beta_0)=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)^\tau\hat{\Sigma}_n^{-1}(\beta_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)+r_n(\beta_0)o_p(1)\xrightarrow{d}Z^\tau\Sigma^{-1}(\beta_0)Z\sim\chi_p^2.$$

This proves part (a).

(b). Part (a) implies immediately that $P(\beta_0\in I_\alpha)=P(\hat{l}_{n,ad}(\beta_0)\le\chi^2_{p,\alpha})\to 1-\alpha$.

(c). It follows from (23), (24) and Lemma A.1 that

$$\tilde{l}_n(\beta_0)=\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)^\tau\hat{\Sigma}_{1n}^{-1}(\beta_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{in}\right)+o_p(1)\xrightarrow{d}Z^\tau\Sigma_1^{-1}(\beta_0)Z. \tag{26}$$

Therefore, $P(\beta_0 \in I_e) = P(\tilde{l}(\beta_0) \le \chi^2_{p,\alpha}) \to P(Z^\tau \Sigma_1^{-1}(\beta_0) Z \le \chi^2_{p,\alpha}) > P(Z^\tau \Sigma^{-1}(\beta_0) Z \le \chi^2_{p,\alpha}) = 1 - \alpha$.

Remark A.1. It follows from (26) that the asymptotic distribution of $\tilde{l}_n(\beta_0)$ is the same as that of $\sum_{l=1}^p w_l \mathcal{X}^2_{1,l}$ where w_l, $l = 1, \ldots, p$, are the eigenvalues of $\Sigma_1^{-1}(\beta_0)\Sigma(\beta_0)$ and the $\chi^2_{1,l}$'s are independent standard chi-square random variables with one degree of freedom.

Proof of Thoerem 3.1. Similar to (23) and (24), it can be shown that

$$\hat{l}_{nk,ad}(\theta_0) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{in}(\theta_0)\right) \hat{\Sigma}_{n0}^{-1}(\theta_0) \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{in}(\theta_0)\right) + o_p(1), \tag{27}$$

$$\hat{\Sigma}_{n0}(\theta_0) \xrightarrow{p} \Sigma_0(\theta_0), \tag{28}$$

where $\Sigma_0(\theta_0) = \Sigma_{10}(\theta_0) - \Sigma_{20}(\theta_0)$. Next we show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{in}(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_0(\theta_0)). \tag{29}$$

Let $\hat{\gamma}_n = \hat{\gamma}_n(\hat{G}_n)$ and $\hat{\theta}_n$ be the subvector of the first k elements of $\hat{\gamma}_n$. Using $\tilde{\boldsymbol{X}}_i^\tau \hat{\gamma}_n = \tilde{\boldsymbol{X}}_{i1}^\tau \hat{\theta}_n + \tilde{\boldsymbol{X}}_{i2}^\tau \tilde{\beta}_{n(k)}(\hat{G}_n)$, it can be verified that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{in}(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\boldsymbol{X}}_{i1}(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_i^\tau \gamma_0) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\boldsymbol{X}}_{i1}\tilde{\boldsymbol{X}}_i^\tau(\gamma_0 - \hat{\gamma}_n) + \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\boldsymbol{X}}_{i1}\tilde{\boldsymbol{X}}_{i1}^\tau(\hat{\theta}_n - \theta_0).\end{aligned} \tag{30}$$

Denote the three terms on the right side of (30) by T_{n1}, T_{n2} and T_{n3} respectively. Then,

$$T_{n2} = -(E[\tilde{\boldsymbol{X}}_{11}\tilde{\boldsymbol{X}}_1^\tau])(E\tilde{\boldsymbol{X}}_1\tilde{\boldsymbol{X}}_1^\tau)^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_i^\tau\gamma_0)\right] + o_p(1), \tag{31}$$

$$\begin{aligned}T_{n3} &= E(\tilde{\boldsymbol{X}}_{11}\tilde{\boldsymbol{X}}_{11}^\tau)\{E(\tilde{\boldsymbol{X}}_{i1}\tilde{\boldsymbol{X}}_{i1}^\tau) - K^\tau P^{-1} K\}^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n (\tilde{\boldsymbol{X}}_{i1} - K^\tau P^{-1}\tilde{\boldsymbol{X}}_{i2})(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_i^\tau\gamma_0)\right] + o_p(1).\end{aligned} \tag{32}$$

It follows from (30), (31) and (32) that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{in}(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_i(Y_{i\hat{G}_n} - \tilde{\boldsymbol{X}}_i^\tau\gamma_0) + o_p(1), \tag{33}$$

where η_i is defined in Theorem 3.1. This, together with the arguments in the proof of Theorem 2 of Lai, Ying and Zheng (1995) leads to (29).

Finally, combining (27), (28) and (29) completes the proof.

Lemma A.2. (i). *If $\psi_{in}(\xi) = W_{in}(\beta)$ and the conditions of Theorems* 2.1 *and* 4.1 *hold, then* $\frac{1}{\sqrt{n}}\sum_{i=1}^n A_{ni}(\beta_0)\xrightarrow{\mathcal{L}}N(0, V_{2,AU}(\beta_0))$, *where* $V_{2,AU}(\beta_0)=\begin{pmatrix} V_1, & 0 \\ 0, & \Sigma(\beta_0)\end{pmatrix}$.

(ii). *If $\psi_{in}(\xi) = u_{in}(\theta)$ and the conditions of Theorems* 3.1 *and* 4.1 *hold, then* $\frac{1}{\sqrt{n}}\sum_{i=1}^n A_{ni}(\theta_0)$ *is asymptotically normal with mean* 0 *and variance-covariance matrix*

$$V_{2,AU}(\theta_0) = \begin{pmatrix} V_1, & 0 \\ 0, & \sigma^2(\theta_0)\end{pmatrix}.$$

Proof. Part (i) is a direct consequence of Lemma A.1 and the following facts: $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i)\xrightarrow{\mathcal{L}}N(0, V_1(\beta_0))$, $\mathrm{Cov}\,(\frac{1}{\sqrt{n}}\sum g(X_i), \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{in}(\beta_0)) \to 0$. Part (ii) can be proved along the same lines.

Proof of Theorem 4.1. The theorem can be proved using Lemma A.2 and the same arguments used in the proof of Theorems 2.1 and 3.1. We omit the details.

Acknowledgements

The authors are grateful to the Editor, and an associate editor and two anonymous referees for their constructive comments that lead to significant improvement of the paper.

References

Adimari, G. (1997). Empirical likelihood type confidence intervals under random censorship. *Ann. Inst. Statist. Math.* **49**, 447-466.

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

Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.* **45**, 621-637.

Chen, S. X. and Cui, H. (2003). An extended empirical likelihood for generalized linear models. *Statist. Sinica.* To appear.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21**, 1166-1181.

Chen, J. H. and Qin J. (1993), Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107-116.

Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *J. Multivariate Anal.* **49**, 24-40.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

DiCiccio, T. J. Hall, P. and Romano, J. P. (1991). Bartlett adjustment for empirical likelihood. *Ann. Statist.* **19**, 1053-1061.

Einmahl, J. H. J. and McKeague, I. W. (1999). Confidence tubes for multiple quantile plots via empirical likelihood. *Ann. Statist.* **27**, 1348-1367.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109-127.

Hollander, M., McKeague, I. W. and Yang, J. (1997). Likelihood ratio-based confidence bands for survival functions. *J. Amer. Statist. Assoc.* **92**, 215-226.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084-2102.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**, 199-218.

Koul, H., Susarla, V. and van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

Lai, T. L., Ying, Z. and Zheng, Z. (1995). Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression. *J. Multivariate Anal.* **52**, 259-279.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301-109.

Li, G. (1995a). Nonparametric likelihood ratio estimation of probabilities for truncated data. *J. Amer. Statist. Assoc.* **90**, 997-1003.

Li, G. (1995b). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. Probab. Lett.* **25**, 95-104.

Li, G., Hollander, M., McKeague, I. W. and Yang, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Ann. Statist.* **24**, 628-640.

Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in nonparametric regression with censored data. *Scand. J. Statist.* **29**, 547-562.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-464.

Murphy, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *J. Amer. Statist. Assoc.* **90**, 1399-1405.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for single functional. *Biometrika* **75**, 237-249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182-1196.

Qin, J. (1999). Empirical likelihood based confidence intervals for mixture proportions. *Ann. Statist.* **27**, 1386-1384.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fits and independence in two-way tables. *J. Amer. Statist. Assoc.* **76**, 221-230.

Srinivasan, C. and Zhou, M. (1991). Linear regression with censoring. *J. Multivariate Anal.* **49**, 179-201.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591-1607.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70**, 865-871.

Wang, Q. H. and Jing, B. Y. (1999). Empirical likelihood for partial linear model with fixed design. *Statist. Probab. Lett.* **41**, 425-433.

Wang, Q. H. and Li, G. (2002). Empirical likelihood semiparametric regression analysis under random censorship. *J. Multivariate Anal.*, in press.

Wang, Q. H. and Wang J.-L. (2001). Inference for the mean difference in the two-sample random censorship model. *J. Multivariate Anal.*, **79**, 295-315.

Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Inference* **44**, 77-94.

Zhang, B. (1996). Confidence intervals for a distribution function in the presence of auxiliary information. *Comput. Statist. Data Anal.* **21**, 327-342.

Zhou, M. (1992). Asymptotic normality of the synthetic estimator for censored survival data. *Ann. Statist.* **20**, 1002-1021.

Zhou, W. and Jing, B.-Y. (2003). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Statist. Sinica*. To appear.

Department of Biostatistics, School of Public Health, Los Angeles, CA 90095-1772.

E-mail: gangli@sunlab.ph.ucla.edu

(Received March 2001; accepted July 2002)