AISM (2008) 60:441–457
DOI 10.1007/s10463-006-0110-4

Empirical likelihood inference for censored median regression with weighted empirical hazard functions

Yichuan Zhao · Song Yang

Received: 12 December 2005 / Revised: 25 August 2006 / Published online: 7 February 2007

Abstract In recent years, median regression models have been shown to be useful for analyzing a variety of censored survival data in clinical trials. For inference on the regression parameter, there have been a variety of semiparametric procedures. However, the accuracy of such procedures in terms of coverage probability can be quite low when the censoring rate is heavy. In this paper, based on weighted empirical hazard functions, we apply an empirical likelihood (EL) ratio method to the median regression model with censoring data and derive the limiting distribution of EL ratio. Confidence region for the regression parameter can then be obtained accordingly. Furthermore, we compared the proposed method with the standard method through extensive simulation studies. The proposed method almost always outperformed the existing method.

Keywords Confidence region · Conditional Kaplan–Meier estimator · Martingale · Counting process · Right censoring · Weighted empirical processes

1 Introduction

In the analysis of survival data, the accelerated failure time (AFT) model is an alternative to the popular Cox proportional hazards model. Its ease of

Y. Zhao (✉)
Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA
e-mail: yzhao@mathstat.gsu.edu

S. Yang
Office of Biostatistics Research, National Heart, Lung, and Blood Institute, 6701 Rockledge Drive, MSC 7938 Bethesda, MD 20892, USA

interpretation makes the AFT model appealing to the practitioners. Some recent work includes Buckley and James (1979), Koul et al. (1981), Lai and Ying (1992), Ritov (1990), Tsiatis (1990), Wei et al. (1990), Yang (1997a,b), among others. When the error has mean zero, the AFT model can be viewed as a mean model.

The median is a simple and meaningful measure for the center of a long-tailed survival function. In biomedical setting, it is frequently of interest to estimate median life length at given covariate levels. Median regression offers an attractive robust alternative to the AFT model. Information about the median life length is immediate once the regression parameters are estimated. For uncensored data, robust regression analysis can be obtained by the least absolute deviations (LADs) method. The work in economic metric research includes Bassett and Koenker (1978), Koenker and Bassett (1978), Powell (1984, 1986) among others.

Ying et al. (1995) studied a median regression model where they allow error to depend on covariate with conditional median zero. When the censoring variables are i.i.d. and independent of the covariate, they use Kapaln–Meier estimator of the censoring distribution and a Koul et al. (1981) type inverse probability weighting. When covariate depends on censoring it requires a nonparametric procedure such as the nearest neighbor method for the censoring distribution. For uncensored data, Jung (1996) obtained an efficient estimating function for the median regression parameters based on quasi-likelihood. Recently, McKeague et al. (2001) applied missing information principle to median regression model and proposed a new estimating function. When the covariate takes values in a finite set, the proposed estimating function is equivalent to the estimating function in Ying et al. (1995). However, in general, the two estimating functions lead to different estimators of the regression parameter.

Yang (1999) proposed certain alternative semiparametric estimators, which were based on some weighted empirical survival and hazard functions. The procedures do not require estimation of the censoring distributions. The assumptions on the censoring distributions and the covariate are quite general: The censoring distributions can be different and covariate dependent, and the covariate need not be discretized. The proposed estimators perform well in the simulation examples. See Yang (1999) for the discussion.

Our approach is based on the empirical likelihood (EL) method. EL method is a powerful nonparametric method. It holds some unique features, such as range respecting, transformation-preserving, asymmetric confidence interval, Bartlett correctability, and better coverage probability for small sample (cf. DiCiccio et al., 1991). Owen (1988, 1990) introduced empirical likelihood confidence regions for the mean of a random vector based on i.i.d. complete data. Since then, the EL has been widely applied to statistical inference. Some related work includes simultaneous confidence band under a variety of setting [see Hollander et al. (1997), Einmahl and McKeague (1999), Li and Van Keilegom (2002), and McKeague and Zhao (2002, 2005, 2006)], linear regression model with right censored data [Li and Wang, 2003; Qin and Jing, 2001], regression analysis of long-term survival rate (Zhao 2005), additive hazard model with right censoring (Zhao and Hsu, 2005), mean residual life function [Zhao and Qin, 2006], weighted EL (Glenn and Zhao, 2006), among others.

More recently, Qin and Tsao (2003) developed EL based confidence region based on estimating equation of Ying et al. (1995). One advantage is that the EL based confidence region is determined by the data set. The theoretical result holds only when censoring is independent of the covariate, or when the covariate is discrete. These constraints limit the application of the proposed method in practice. Moreover, their simulation results show that the methods of Ying et al. (1995) and Qin and Tsao (2003) had some undercoverage problems, sometimes severely for small sample and heavy censoring.

In order to overcome the limitation of their methods, we use the EL approach based on the estimating equation from Yang (1999). Specifically, we consider the following median regression model for censored data. Let T_i $(i = 1, \ldots, n)$ be the response of interest. Let $Z_i = (1, X_i')'$, where X_i is a $p \times 1$ vector, be the corresponding $p + 1$ dimensional covariate vector. Then the median regression model is given by

$$T_i = \beta' Z_i + \epsilon_i, \tag{1}$$

where β is a $(p+1) \times 1$ vector of unknown regression parameter. The median of ϵ_i is zero. We observe (Y_i, Δ_i), where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. The censoring variable C_i is assumed to be conditionally independent of T_i given the covariate Z_i for $1 \leq i \leq n$. We find EL and adjusted EL confidence regions for the unknown regression parameter. The simulation results demonstrate the proposed EL method is more accurate than existing methods in terms of coverage probability for small sample size.

The rest of the article is organized as follows. The proposed unadjusted EL and adjusted EL confidence regions and main asymptotic result are presented in Sect. 2. In Sect. 3, we conduct an extensive simulation study using fixed design and random design. Proofs are contained in the Appendix.

2 Main results

2.1 Preliminaries

We consider the median regression model (1) with nonrandom covariate and homogeneous errors. Suppose ϵ_is are i.i.d. with common cdf F whose median is 0. We assume the censoring variables C_is are identically distributed with cdf G. Let $Z = (Z_1, \ldots, Z_n)'$ be the $n \times (p+1)$ covariate matrix, with (i, j) element $Z_{ij}, i = 1, \ldots, n, j = 1, \ldots, p+1$, where $Z_{i1} = 1$. For $-\infty < t < \infty$ and any fixed $k \times 1$ vector b, let $e_i(b) = Y_i - b' Z_i$ and

$$K_j(t; b) = \sum_{i=1}^{n} Z_{ij} I(e_i(b) \geq t), \;\; j = 1, \ldots, p+1. \tag{2}$$

Yang (1999) defined the weighted empirical hazard functions,

$$\hat{\Lambda}_j(t;b) = \sum_{e_i(b)\le t} \frac{Z_{ij}\Delta_i}{K_j(e_i(b);b)},\ \ j=1,\ldots,p+1, \tag{3}$$

and proposed the following estimating equation

$$\hat{\Lambda}_j(0;b) = \log\ 2,\ \ j=1,\ldots,p+1. \tag{4}$$

Under mild conditions the estimating equation has a unique solution $\hat{\beta}$.
Define

$$\Gamma_{nj}(t;b) = \frac{1}{n}\sum_{i=1}^{n} Z_{ij}\bar{G}_-(t+Z_i'b),$$

$$\Gamma_{njl}(t;b) = \frac{1}{n}\sum_{i=1}^{n} Z_{ij}Z_{il}\bar{G}_-(t+Z_i'b),$$

where $\bar{D}=1-D$ for any distribution function D, $A_-(t)$ denotes the left limit version of the function $A(t)$. Let β be the true value of regression parameter in model (1). As in Yang (1999), to derive the asymptotic normality of $\hat{\beta}$, we need the following conditions:

1. $T_1,\ldots,T_n$ and $C_1,\ldots,C_n$ are independent and Z_{ij} are nonrandom constants.
2. The covariate vector Z is nonnegative and bounded, i.e., $||Z||\le M$ for some positive constant M, where $||\cdot||$ is the Euclidean norm.
3. The limits $\Gamma_j(t)=\lim_n \Gamma_{nj}(t;\beta), j=1,\ldots,k$, and $\Gamma_{jl}(t)=\lim_n \Gamma_{njl}(t;\beta), j,l=1,\ldots,k$, exist, and $\inf_j \Gamma_j(c_0)>0$ for some $c_0>0$.
4. $\inf_j \liminf_n \Gamma_{nj}(0;b) > 0$ for b in a compact neighborhood N of β. G has Lebesgue density g. $\sup_t |g(t)| \le B$ for some $B>0$ and $E|T_1\wedge C_1|^r<\infty$ for some $r>0$. F has Lebesgue density f. f' is uniformly continuous and integrable. $\int_{-\infty}^{c_0} w(t,a_n)\,\mathrm{d}t = o(1)$ as $a_n\to 0$, where

$$w(t,a_n) = \sup_{|s|<a_n} (|f'(t+s)-f'(t)| + |f(t+s)-f(t)|).$$

5. β is in the interior of the region N. For some $\alpha,\epsilon>0$, $\inf_j \liminf_n \inf_{b\in N,||b||>\alpha}$ $n^{1/2-\epsilon}|S_{nj}(0;\beta+b)-1/2|>0$, where $S_{nj}(0;b)$ is defined in Yang (1999).
6. The limits

$$\Phi_{jl} = \frac{1}{2}\lim_n \int_{-\infty}^{0} \frac{\Gamma'_{njl}(t;\beta)}{\Gamma_{nj}(t;\beta)}\,\mathrm{d}\lambda(t),\ \ j,l=1,\ldots,p+1, \tag{5}$$

exist and the $(p+1)\times(p+1)$ matrix $\Phi = (\Phi_{jl})$ is nonsingular, where $\lambda = f/\bar{F}$ is the hazard rate function corresponding to f.

Remark The above regularity conditions are commonly used in survival analysis, see Ritov (1990) and Ying (1993) for discussion. Condition 1 is the most basic model assumption. The bounded assumption in condition 2 is standard for covariate in survival analysis, see Ying et al. (1995) and p. 421 of Qin and Tsao (2003). Since covariate Z is bounded, the nonnegativity of Z in condition 2 is satisfied by a proper transformation if necessary. Condition 3 guarantees the limit of $S_{nj}(0;\beta)$ in condition 5 is well defined and equals 1/2. Condition 4 is needed to approximate $\hat{\Lambda}_j$ by their deterministic counterparts (cf. Ying (1993)). These assumptions are satisfied by the common distributions in survival analysis. Condition 4 guarantees that $\hat{\beta}$ is asymptotically uniquely defined and consistent. It can be verified that condition 5 is satisfied for the k-sample problem if f is positive in a neighborhood of 0. Condition 6 ensures that asymptotically the estimators are well defined. This condition is similar to the requirement that $Z'Z$ be invertible in the ordinary regression with no censoring. See pp. 142–143 of Yang (1999) for the discussion.

When the covariate is random, the above regularity conditions need to be modified accordingly, and a set of regularity conditions is given in the remark of Yang (1999).

Under above regularity conditions, Yang (1999) showed that

$$n^{1/2}(\hat{\beta}-\beta) \xrightarrow{\mathcal{D}} N(0, \Phi^{-1}V\Phi^{-1}), \tag{6}$$

where (j,l) element of V is

$$v_{jl} = \int_{-\infty}^{0} \frac{\Gamma_{jl}}{\Gamma_j\Gamma_l\bar{F}}\,\mathrm{d}\Lambda, \quad j,l = 1,\ldots,k, \tag{7}$$

and $\Lambda(t) = \int_{-\infty}^{t}\lambda(s)\mathrm{d}s$.

The asymptotic covariance matrix of $\hat{\beta}$ involves the error density functions and usually difficult to estimate. Resampling method such as that of Parzen et al. (1994), can be used to obtain the estimated covariance matrix and then confidence region for β.

Let $A_n(b)$ be the $(p+1)\times 1$ vector with jth component $\sqrt{n}(\hat{\Lambda}_j(0;b) - \log 2)$. In the Appendix of Yang (1999), under some regularity conditions, $A_n(\beta)$ has a limiting normal distribution, with mean zero and covariance matrix V. Let

$$K_{jl}(t;b) = \sum_{i=1}^{n} Z_{ij}Z_{il}I(e_i(b)\geq t), \;\; j,l = 1,\ldots,p+1. \tag{8}$$

The covariance matrix V can be consistently estimated by $\hat{V}$, whose (j, l) element is

$$\hat{v}_{jl} = n \int_{-\infty}^{0} \frac{K_{jl}(t; \hat{\beta})}{K_j(t; \hat{\beta}) K_l(t; \hat{\beta})} \hat{\Lambda}_1(\mathrm{d}t; \hat{\beta}). \tag{9}$$

Thus, using the test-based approach, an asymptotic $100(1 - \alpha)\%$ confidence region for β is given by

$$\mathcal{R}_1 = \left\{ b : A_n'(b) \hat{V}^{-1} A_n(b) \leq \chi^2_{p+1}(\alpha) \right\},$$

where $\chi^2_{p+1}(\alpha)$ is the upper α-quantile of the chi-squared distribution with degrees of freedom $p + 1$.

2.2 EL confidence region

Now we introduce the EL approach based on the estimating equations in Yang (1999). The estimating equation can be written as

$$\sum_{i=1}^{n} \left(\frac{Z_{ij} \Delta_i I(e_i(b) \leq 0)}{K_j(e_i(b); b)} - \frac{\log 2}{n} \right) = 0, \;\; j = 1, \ldots, p + 1. \tag{10}$$

For $1 \leq i \leq n$, we define the $(p + 1) \times 1$ vector W_{ni} with jth element

$$W_{ni}(j) = n \left(\frac{Z_{ij} \Delta_i I(e_i(\beta) \leq 0)}{K_j(e_i(\beta); \beta)} - \frac{\log 2}{n} \right), \;\; j = 1, \ldots, p + 1.$$

Then, the EL is given by

$$L(\beta) = \sup \left\{ \prod_{i=1}^{n} p_i : \sum p_i = 1, \sum_{i=1}^{n} p_i W_{ni} = 0, \;\; p_i \geq 0, i = 1, \ldots, n \right\}.$$

Let $p = (p_1, \ldots, p_n)$ be a probability vector, i.e., $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. Note that $\prod_{i=1}^{n} p_i$ attains its maximum at $p_i = 1/n$. Thus, the EL ratio at the true value β is defined by

$$R(\beta) = \sup \left\{ \prod_{i=1}^{n} n p_i : \sum p_i = 1, \sum_{i=1}^{n} p_i W_{ni} = 0, \;\; p_i \geq 0, i = 1, \ldots, n \right\}.$$

By the method of Lagrange multipliers, we know that $R(\beta)$ is maximized when

$$p_i = \frac{1}{n} \{1 + \lambda' W_{ni}\}^{-1}, \;\; i = 1, \ldots, n,$$

where $\lambda = (\lambda_1, \ldots, \lambda_{p+1})'$ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{W_{ni}}{1+\lambda' W_{ni}} = 0. \tag{11}$$

The value of λ may be found by numerical search (e.g., Newton–Raphson method), see the discussion in Hall and La Scala (1990). Thus, combining above equalities we have the corresponding empirical log-likelihood

$$\hat{l}(\beta) = -2\log R(\beta) = -2\log\prod_{i=1}^{n}(np_i) = 2\sum_{i=1}^{n}\log\{1+\lambda' W_{ni}\}, \tag{12}$$

where λ satisfies Eq. (11).
Let

$$U_j(t;b) = \lim_{n\to\infty} K_j(t;b)/n, \quad j = 1,\ldots,p+1. \tag{13}$$

For $1 \le i \le n$, we define the $(p+1)\times 1$ vector W_i with jth element

$$W_i(j) = \frac{Z_{ij}\Delta_i I(e_i(\beta)\le 0)}{U_j(e_i(\beta);\beta)} - \log 2, \quad j = 1,\ldots,p+1.$$

Let

$$V_1 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} W_i W_i'.$$

In Lemma 1 of the Appendix, we show that $V_1 = V - (\log 2)^2$. Now we state our main result and explain how it can be used to construct confidence region for β.

Theorem 1 *Under the above conditions* 1–6*, the EL statistic* $-2\log R(\beta)$ *converges in distribution to* $r_1\chi^2_{1,1} + \cdots + r_{p+1}\chi^2_{p+1,1}$*, where* $\chi^2_{1,1}, \ldots, \chi^2_{p+1,1}$ *are independent chi-square random variables with 1 degree of freedom and* $r_1, \ldots, r_{p+1}$ *are the eigenvalues of* $V_1^{-1}V$*.*

Theorem 1 will be proved in the Appendix. We note that the limiting distribution of the EL ratio is a weighted sum of i.i.d. χ^2_1s instead of the standard χ^2_{p+1} distribution. This is due to the fact that the W_{ni}'s are dependent. Similar phenomenon has occurred in various contexts, such as Qin and Jing (2001), Wang and Rao (2001), Wang and Wang (2001), and Li and Wang (2003), among others.

Although the limiting distribution has the nonstandard weighted sum expression, the weights involved can be readily estimated so that the above theorem can be used in parameter inference. For $1 \le i \le n$, let $\hat{W}_{ni}$ be the $(p+1) \times 1$ vector with jth element

$$\hat{W}_{ni}(j) = n\left(\frac{Z_{ij}\Delta_i I(e_i(\hat{\beta}) \le 0)}{K_j(e_i(\hat{\beta}); \hat{\beta})} - \frac{\log 2}{n}\right), \quad j = 1, \ldots, p+1,$$

and define

$$\hat{V}_1 = n^{-1}\sum_{i=1}^{n} \hat{W}_{ni}\hat{W}'_{ni}.$$

From Lemma 1 (i), V_1 is consistently estimated by $\hat{V}_1$. Hence, the r_is can be estimated by the $\hat{r}_i$'s which are the eigenvalues of $\hat{V}_1^{-1}\hat{V}$. An asymptotic $100(1-\alpha)\%$ confidence region for β is given by

$$\mathcal{R}_2 = \{b : -2\log R(b) \le c(\alpha)\},$$

where $c(\alpha)$ is the upper α-quantile of the distribution of $\hat{r}_1\chi^2_{1,1} + \cdots + \hat{r}_{p+1}\chi^2_{p+1,1}$ and can be obtained by simulation method.

Alternatively, the above EL approach can be adjusted to avoid the weighted sum expression. Let $\rho(\beta) = (p+1)/\mathrm{tr}\{V_1^{-1}(\beta)V(\beta)\}$ with $\mathrm{tr}(\cdot)$ denoting the trace vector, i.e., the trace of a matrix. Then, following Rao and Scott (1981), the distribution of $\rho(\beta)(r_1\chi^2_{1,1} + \cdots + r_{p+1}\chi^2_{p+1,1})$ may be approximated by χ^2_{p+1}. This implies that the asymptotic distribution of the Rao–Scott adjusted empirical likelihood ratio, $\tilde{l}_{ad}(\beta) = \hat{\rho}(\beta)\hat{l}(\beta)$, may be approximated by χ^2_{p+1}, where the adjustment factor $\hat{\rho}(\beta)$ is $\rho(\beta)$ with $V_1(\beta)$ and $V(\beta)$ replaced by $\hat{V}_1(\beta)$ and $\hat{V}(\beta)$, respectively.

The adjusted EL approach was proposed by Wang and Rao (2001, 2002) and Li and Wang (2003), among others. We define an adjusted EL ratio, by modifying $\rho(\beta)$ in $\tilde{l}_{ad}(\beta)$, whose asymptotic distribution is exactly a standard chi-square distribution with $p+1$ degrees of freedom, i.e., χ^2_{p+1}. Noting that

$$\hat{\rho}(\beta) = \frac{\mathrm{tr}\left\{\hat{V}^{-1}(\beta)\hat{V}(\beta)\right\}}{\mathrm{tr}\left\{\hat{V}_1^{-1}(\beta)\hat{V}(\beta)\right\}},$$

we define $\hat{r}(\beta)$ to be $\hat{\rho}(\beta)$ with $\hat{V}(\beta)$ replaced by $\hat{S}(\beta) = \{\sum_{i=1}^{n} W_{ni}(\beta)/n\} \times \{\sum_{i=1}^{n} W_{ni}(\beta)/n\}'$. That is,

$$\hat{r}(\beta) = \frac{\mathrm{tr}\{\hat{V}^{-1}(\beta)\hat{S}(\beta)\}}{\mathrm{tr}\{\hat{V}_1^{-1}(\beta)\hat{S}(\beta)\}}.$$

We define an adjusted EL ratio by

$$\hat{l}_{ad}(\beta) = \hat{r}(\beta)\hat{l}(\beta).$$

Theorem 2 *Under the above conditions* 1–6*, the EL statistic* $\hat{l}_{ad}(\beta)$ *converges in distribution to* χ^2_{p+1}.

Based on Theorem 2, an asymptotic $100(1-\alpha)\%$ confidence region for β is given by

$$\mathcal{R}_3 = \{b : \hat{l}_{ad}(b) \leq \chi^2_{p+1}(\alpha)\},$$

where $\chi^2_{p+1}(\alpha)$ is define as before.

3 Simulation study

Various simulation studies have been conducted to assess the behavior of the EL approach for the median regression model (1). We compared the performance of the proposed EL confidence region with normal approximation (NA) confidence region in terms of coverage probability.

We generate data $(Y_i, \Delta_i, Z_i), i = 1, \ldots, n$, where $Z_i = (1, X_i)'$, n is sample size, and obtain model (1). Let $U_i, i = 1, \ldots, n$ be i.i.d. Uniform (0, 1), let $N_i, M_i, i = 1, \ldots, n$ be the i.i.d. standard normal. Also let $u_i, i = 1, \ldots, n$ be a fixed realized sample from Uniform (0, 1). Let c be a constant which controls the censoring rate (CR). In one such study, the following models are considered, which represent fixed design and random design.
Model 1: $X_i = u_i, T_i = X_i + 0.5N_i$, and $C_i = c_i + 0.5M_i$
Model 2: $X_i = i/n, T_i = X_i + 0.5N_i$, and $C_i = c + i/n + 0.5M_i$
Model 3: $X_i = U_i, T_i = X_i + 0.5N_i$, and $C_i = c + 0.5M_i$
Model 4: $X_i = U_i, T_i = X_i + 0.5N_i$, and $C_i = c + X_i + 0.5M_i$

The true parameter β is $(0, 1)'$. We take 0.90, 0.95, and 0.99 as the nominal confidence level $1-\alpha$, respectively. We obtain 20, 40, and 60% censoring rates, respectively, which represent light censoring, medium censoring, and heavy censoring. The sample size n is chosen to be 40, 60, 80, and 100, respectively. The coverage probabilities of the normal approximation based method and the empirical likelihood method are estimated from 2,000 simulated data sets. The simulation results for models are reported in Tables 1–4 respectively.

From the tables, we make the following observations.

1. At each nominal level, the coverage accuracies for EL and NA methods in all models including fixed design and random design decrease as CRs increase, and increase when sample size increases.

2. The coverage probabilities for the NA method and EL method are consistently lower than the nominal level for small sample ($n = 40$).

Table 1 Coverage probabilities for the regression parameter in model 1

		$1-\alpha=0.90$		$1-\alpha=0.95$		$1-\alpha=0.99$	
CR(%)	n	NA	EL	NA	EL	NA	EL
	40	0.881	0.898	0.936	0.941	0.978	0.976
20	60	0.884	0.890	0.936	0.936	0.978	0.981
	80	0.892	0.904	0.946	0.951	0.982	0.985
	100	0.906	0.903	0.949	0.950	0.986	0.990
	40	0.885	0.891	0.935	0.940	0.975	0.981
40	60	0.890	0.894	0.939	0.946	0.983	0.989
	80	0.904	0.905	0.944	0.954	0.982	0.986
	100	0.896	0.904	0.947	0.954	0.984	0.988
	40	0.876	0.896	0.925	0.945	0.965	0.975
60	60	0.885	0.908	0.929	0.955	0.974	0.988
	80	0.883	0.905	0.934	0.956	0.979	0.990
	100	0.879	0.898	0.934	0.956	0.983	0.992

Table 2 Coverage probabilities for the regression parameter in model 2

		$1-\alpha=0.90$		$1-\alpha=0.95$		$1-\alpha=0.99$	
CR(%)	n	NA	EL	NA	EL	NA	EL
	40	0.887	0.895	0.939	0.936	0.977	0.976
20	60	0.891	0.906	0.942	0.945	0.984	0.984
	80	0.889	0.899	0.947	0.954	0.986	0.991
	100	0.901	0.898	0.946	0.951	0.987	0.988
	40	0.880	0.893	0.924	0.938	0.974	0.984
40	60	0.893	0.903	0.945	0.955	0.985	0.989
	80	0.891	0.908	0.940	0.949	0.985	0.990
	100	0.891	0.897	0.942	0.952	0.985	0.988
	40	0.873	0.896	0.914	0.938	0.961	0.977
60	60	0.874	0.907	0.932	0.956	0.973	0.987
	80	0.879	0.891	0.934	0.949	0.978	0.989
	100	0.885	0.911	0.932	0.956	0.981	0.988

3. The EL outperforms the NA method in all models. In particular, under heavy censoring rate (CR = 60%), the EL confidence region has more accurate coverage probabilities than the NA based confidence region. At other censoring rates, the EL performs better than the NA method in almost all cases. As sample size increases the advantage of EL disappears gradually as expected.

From Tables 1–4, we find that the normal approximation based method does not always work well for sample. One reason is that the NA based confidence region needs to estimate V [cf. (9) of Sect. 2.1]. The variance estimates are not very stable and may contain values outside their ranges.

We have also conducted a simulation with adjusted EL method. When the sample size is small (n = 40, 60), the adjusted EL results in lower coverage probabilities than nominal level. As the sample size increases, the adjusted EL begins to perform comparably with the unadjusted EL. For sample size $n = 500$,

Table 3 Coverage probabilities for the regression parameter in model 3

		$1-\alpha=0.90$		$1-\alpha=0.95$		$1-\alpha=0.99$	
CR(%)	n	NA	EL	NA	EL	NA	EL
	40	0.888	0.892	0.939	0.944	0.977	0.980
20	60	0.882	0.890	0.936	0.944	0.983	0.985
	80	0.894	0.895	0.939	0.943	0.984	0.985
	100	0.890	0.896	0.940	0.946	0.986	0.990
	40	0.870	0.879	0.925	0.938	0.982	0.984
40	60	0.880	0.889	0.933	0.941	0.977	0.989
	80	0.894	0.907	0.946	0.955	0.986	0.989
	100	0.893	0.905	0.944	0.953	0.984	0.989
	40	0.865	0.883	0.921	0.938	0.967	0.978
60	60	0.882	0.905	0.926	0.948	0.968	0.984
	80	0.878	0.917	0.937	0.963	0.979	0.992
	100	0.890	0.910	0.938	0.961	0.985	0.991

Table 4 Coverage probabilities for the regression parameter in model 4

		$1-\alpha=0.90$		$1-\alpha=0.95$		$1-\alpha=0.99$	
CR(%)	n	NA	EL	NA	EL	NA	EL
	40	0.883	0.886	0.941	0.940	0.976	0.976
20	60	0.890	0.889	0.935	0.940	0.980	0.983
	80	0.892	0.907	0.944	0.952	0.985	0.986
	100	0.893	0.896	0.938	0.940	0.982	0.987
	40	0.880	0.889	0.928	0.943	0.979	0.982
40	60	0.890	0.905	0.943	0.947	0.980	0.983
	80	0.891	0.894	0.941	0.948	0.986	0.988
	100	0.889	0.900	0.946	0.947	0.987	0.991
	40	0.868	0.890	0.917	0.939	0.965	0.980
60	60	0.871	0.907	0.927	0.951	0.974	0.988
	80	0.882	0.913	0.937	0.961	0.979	0.989
	100	0.885	0.910	0.934	0.956	0.981	0.993

the unadjusted EL is more conservative and the adjusted EL is more accurate than the unadjusted EL. We omit the results here. Similar phenomenon has been noticed elsewhere (Li and Wang 2003).

From the simulation results, we see that the improvement of our method over Yang (1999) is similar to that of Qin and Tsao (2003) over Ying et al. (1995). Their simulation results show that the methods of Ying et al. (1995) and Qin and Tsao (2003) had some undercoverage problems, sometimes severely for small sample and heavy censoring. In comparison, our proposed EL method has much improvement regarding the undercoverage issue in the current EL methods.

In summary, our simulation study shows that the proposed EL method gives competitive coverage probabilities and suggests that the empirical likelihood improves the coverage in this case.

Acknowledgment The research of Yichuan Zhao is partially supported by a grant from the National Security Agency and Faculty Mentored Grant, Georgia State University. The research of Song Yang is partially supported by grants from the National Science Foundation and the Advanced Research Program of Texas. The authors would like to thank the referees for their helpful comments which improved the presentation of the paper greatly.

Appendix: Proofs of Theorems 1 and 2

In order to prove Theorem 1, we first prove the following lemma.

Lemma 1 *Under the conditions of Theorem 1, we have*
(i) $\sum_{i=1}^{n} W_{ni} W'_{ni}/n \stackrel{\mathcal{P}}{\to} V - (\log\ 2)^2$, *(ii)* $\hat{V}_1 \stackrel{\mathcal{P}}{\to} V - (\log 2)^2$.

Proof For $1 \le j, l \le p+1$, let $c_0 = \log\ 2$ and

$$A_n(t; b) = \frac{K_j(t; b)}{n} \frac{K_l(t; b)}{n},$$

$$B_n(t; b) = \frac{\sum_i Z_{ij} Z_{il} \Delta_i I(e_i(b) \le t)}{n}.$$

Then the (j, l) entry of $\hat{V}_1$ is

$$\begin{aligned}
&\sum_i \frac{n Z_{ij} Z_{il} \Delta_i I(e_i(\hat{\beta}) \le 0)}{K_j(e_i(\hat{\beta}); \hat{\beta}) K_l(e_i(\hat{\beta}); \hat{\beta})} \\
&- c_0 \left(\sum_i \frac{Z_{ij} \Delta_i I(e_i(\hat{\beta}) \le 0)}{K_j(e_i(\hat{\beta}); \hat{\beta})} + \sum_i \frac{Z_{il} \Delta_i I(e_i(\hat{\beta}) \le 0)}{K_l(e_i(\hat{\beta}); \hat{\beta})} \right) + c_0^2 \\
&= \int_{-\infty}^{0} \frac{\mathrm{d}B_n(t; \hat{\beta})}{A_n(t; \hat{\beta})} - c_0 \left(\hat{\Lambda}_j(0, \hat{\beta}) + \hat{\Lambda}_l(0, \hat{\beta}) \right) + c_0^2 \\
&= I + II + c_0^2,
\end{aligned}$$

say.

Similarly to Theorems 1 and 3 in Lai and Ying (1988) or Theorem 2 of Yang (1997), we have, for any $c, \epsilon > 0$, $0 < r < 1$, w. p. 1,

$$\begin{aligned}
&\sup_{\|b\ | < c, t \le 0} |Q_n(t; b) - EQ_n(t; b)| = o(n^{-1/2+\varepsilon}), \\
&\sup_{\|b - b'\| < cn^{-r},\ t \le 0} |Q_n(t; b) - EQ_n(t; b) - Q_n(t; b') + EQ_n(t; b')| \\
&\quad = o(n^{-1/2 - r/2 + \varepsilon}),
\end{aligned}$$

where Q_{nj} is either $A_n(t; b)$ or $B_n(t; b)$.

From these and the $\sqrt{n}$- boundedness of $\hat{\beta}$ in probability as shown in Yang (1999), we obtain that, uniformly in $t \leq 0$, both $A_n(t;\hat{\beta})$ and $B_n(t;\hat{\beta})$ converge in probability to their respective limits $A(t)$, $B(t)$, say. We can also check that $A(t)$ is continuous and the total variations of $B_n(t;\hat{\beta})$ are bounded uniformly in n. Thus, similarly to Lemma A2 of Yang and Prentice (2005), I converges in probability to its limit which can be verified to be v_{jl} in (7). For II it is $-2c_0^2$. Thus the (j,l) entry of $\hat{V}_1$ converges in probability to $v_{jl} - c_0^2$. Thus, (ii) follows. Similar and simpler arguments show that the average of $W_{ni}W_{ni}'$s converges to $V - c_0^2$ and (i) follows. □

Proof of Theorem 1 By condition 2 and the martingale representation of W_{ni}, we can prove $\max_{1\leq i\leq n} ||W_{ni} - W_i|| = o_P(n^{1/2})$. Also we have $\max_{1\leq i\leq n} ||W_i|| = o_P(n^{1/2})$ from $E||W_i||^2 < \infty$ [cf. Owen, 1990]. Thus,

$$\max_{1\leq i\leq n} ||W_{ni}|| \leq \max_{1\leq i\leq n} ||W_{ni} - W_i|| + \max_{1\leq i\leq n} ||W_i|| \\ = o_P(n^{1/2}). \tag{14}$$

Let $\lambda = \rho\theta$ where $\rho \geq 0$ and $||\theta|| = 1$. Recall $\hat{V}_{1n} = V_1 + o_1(1)$ (cf. Lemma 1). We have

$$\theta'\hat{V}_{1n}\theta = \theta' V_1 \theta + o_P(1).$$

Let $\sigma_1 > 0$ be the smallest eigenvalue of V_1. Then, we have

$$\theta'\hat{V}_1\theta \geq \sigma_1/2 + o_P(1). \tag{15}$$

Let e_j be the unit vector in the jth coordinate direction. From the Appendix of Yang (1999),

$$\left| \frac{1}{n} \sum_{j=1}^{p+1} e_j' \sum_{i=1}^{n} W_{ni} \right| = O_P(n^{-1/2}). \tag{16}$$

Then, it follows from (11), (14), (15), (16), and the argument used in Owen (1990) that

$$||\lambda|| = O_P(n^{-1/2}). \tag{17}$$

Applying Taylor's expansion to (12), we have

$$-2\log R(\beta) = 2\sum_{i=1}^{n} \left(\lambda' W_{ni} - \frac{1}{2}(\lambda' W_{ni})^2 \right) + r_n, \tag{18}$$

where

$$|r_n| \leq C\sum_{i=1}^{n}|\lambda' W_{ni}|^3 \ \text{ in } \quad \text{probability.}$$

Hence, by (14), (17)

$$|r_n| \leq Cn||\lambda||^3\left(\max_{1\leq i\leq n}||W_{ni}||\right)^3 = o_P(1). \tag{19}$$

Note that

$$\begin{aligned} 0 = \frac{1}{n}\sum_{i=1}^{n}\frac{W_{ni}}{1+\lambda' W_{ni}} &= \frac{1}{n}\sum_{i=1}^{n}W_{ni}\left(1-\lambda' W_{ni}+\frac{(\lambda' W_{ni})^2}{1+\lambda' W_{ni}}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}W_{ni}-\left(\frac{1}{n}\sum_{i=1}^{n}W_{ni}W'_{ni}\right)\lambda\\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{W_{ni}(\lambda' W_{ni})^2}{1+\lambda' W_{ni}}. \end{aligned} \tag{20}$$

By (16), (17), (20), and Lemma 1, it follows that

$$\lambda = \left(\sum_{i=1}^{n}W_{ni}W'_{ni}\right)^{-1}\sum_{i=1}^{n}W_{ni}+o_P(n^{-1/2}). \tag{21}$$

By (20), we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\frac{\lambda' W_{ni}}{1+\lambda' W_{ni}}\\ &= \sum_{i=1}^{n}(\lambda' W_{ni})-\sum_{i=1}^{n}(\lambda' W_{ni})^2+\sum_{i=1}^{n}\frac{(\lambda' W_{ni})^3}{1+\lambda' W_{ni}}. \end{aligned} \tag{22}$$

Similarly as before by (14), (17),

$$\sum_{i=1}^{n}\frac{(\lambda' W_{ni})^3}{1+\lambda' W_{ni}} = o_P(1). \tag{23}$$

Combining (22) and (23) we have

$$\sum_{i=1}^{n}(\lambda' W_{ni})^2 = \sum_{i=1}^{n}\lambda' W_{ni} + o_P(1). \tag{24}$$

By (18), (19), (21), (24) and Lemma 1, we have

$$\begin{aligned} -2\log R(\beta) &= \sum_{i=1}^{n}\lambda' W_{ni} + o_P(1) \\ &= \left(n^{-1/2}\sum_{i=1}^{n} W_{ni}\right)' \left(n^{-1}\sum_{i=1}^{n} W_{ni}W'_{ni}\right)^{-1} \left(n^{-1/2}\sum_{i=1}^{n} W_{ni}\right) + o_P(1) \\ &= \left(V^{-1/2}n^{-1/2}\sum_{i=1}^{n} W_{ni}\right)' (V^{1/2}V_1^{-1}V^{1/2}) \left(V^{-1/2}n^{-1/2}\sum_{i=1}^{n} W_{ni}\right) + o_P(1). \end{aligned}$$

By the proof of Theorem 1 in Yang (1999), we have $V^{-1/2}(n^{-1/2}\sum_{i=1}^{n} W_{ni}) \xrightarrow{\mathcal{D}} N(0, I_{p+1})$. Because $V^{1/2}V_1^{-1}V^{1/2}$ and $V_1^{-1}V$ have the same eigenvalues. Using Lemma 3 of Qin and Jing (2001) to re-express the limiting distribution of $-2\log R(\beta)$ as a weighted sum of independent χ_1^2 distribution, we complete the proof of Theorem 1. □

Proof of Theorem 2 Recall the definition of $\hat{l}_{ad}(\beta)$. It follows that, by (18),

$$\hat{l}_{ad}(\beta) = \left(n^{-1/2}\sum_{i=1}^{n} W_{ni}\right)' \hat{V}^{-1}\left(n^{-1/2}\sum_{i=1}^{n} W_{ni}\right) + o_P(1).$$

We can show that $\hat{V} \xrightarrow{\mathcal{P}} V$. Using Lemma 3 of Qin and Jing (2001), we complete the proof of Theorem 2. □

References

Bassett, G., Koenker, R. (1978). Asymptotic theory of least absolute regression. *Journal of the American Statistical Association, 73,* 618–622.

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

DiCiccio, T. J., Hall, P., Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annnals of Statistics, 19,* 1053–1061.

Einmahl, J. H., McKeague, I. W. (1999). Confidence tubes for multiple quantile plots via empirical likelihood. *The Annals of Statistics, 27,* 1348–1367.

Glenn, N. L., Zhao, Y. (2006). Weighted empirical likelihood estimates and their robustness properties. *Computational Statistics and Data Analysis,* in press.

Hall, P., La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review, 58,* 109–127.

Hollander, M., McKeague, I. W., Yang, J. (1997). Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association, 92,* 215–226.

Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association, 91,* 251–257.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica, 46,* 33–50.
Koul, H., Susarla, V., Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *The Annals of Statistics, 9,* 1276–1288.
Lai, T. L., Ying, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *Journal of Multivariate Analysis, 27,* 334–358.
Lai, T. L., Ying, Z. (1992). Linear rank statistics in regression analysis with censored or truncated data. *Journal of Multivariate Analysis, 40,* 13–45.
Li, G., Van Keilegom, I. (2002). Likelihood ratio confidence bands in nonparametric regression with censored data. *Scandinavian Journal of Statistics, 29,* 547–562.
Li, G., Wang, Q. H. (2003). Empirical likelihood regression analysis for right censored data. *Statistica Sinica, 13,* 51–68.
McKeague, I. W., Subramanian, S., Sun, Y. (2001). Median regression and the missing information principle. *Journal of Nonparametric Statistics, 13,* 709–727.
McKeague, I. W., Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics and Probability Letters, 60,* 405–415.
McKeague, I. W., Zhao,Y. (2005). Comparing distribution function vial empirical likelihood. *The International Journal of Biostatistics, 5,* 1–20.
McKeague, I. W., Zhao, Y. (2006). Width-scaled confidence bands for survival functions. *Statistics and Probability Letters, 76,* 327–339.
Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika, 75,* 237–249.
Owen, A. (1990). Empirical likelihood and confidence regions. *The Annals of Statistics, 18,* 90–120.
Parzen, M. I., Wei, L. J., Ying, Z. (1994). A resampling method based on pivotal functions. *Biometrika, 81,* 341–350.
Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics, 25,* 303–325.
Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics, 32,* 143–155.
Qin, G., Jing, B. Y. (2001). Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics, 28,* 661–673.
Qin, G., Tsao, M. (2003). Empirical likelihood inference for median regression models for censored survival data. *Journal of Multivariate Analysis, 85,* 416–430.
Rao, J. N. K., Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association, 76,* 221–230.
Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics, 18,* 303–328.
Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics, 18,* 354–372.
Wang, Q. H., Jing, B. Y. (1999). Empirical likelihood for partial linear models with fixed designs. *Statistics and Probability Letters, 41,* 425–433.
Wang, Q. H., Rao, J. N. K. (2001). Empirical likelihood for linear regression models under imputation for missing responses. *The Canadian Journal of Statistics, 29,* 597–608.
Wang, Q. H., Rao, J. N. K. (2002). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika, 89,* 345-358.
Wang, Q. H., Wang, J. L. (2001). Inference for the mean difference in the two-sample random censorship model. *Journal of Multivariate Analysis, 79,* 295–315.
Wei, L. J., Ying, Z., Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika, 19,* 845–851.
Yang, S. (1997a). A generalization of the product-limit estimator with an application to censored regression. *The Annals of Statistics, 25,* 1088-1108.
Yang, S. (1997b). Extended weighted log-rank estimating functions in censored regression. *Journal of the American Statistical Association, 92,* 977–984.
Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association, 94,* 137–145.
Yang, S., Prentice, R. L. (2005). Semiparametric analysis of short term and long term relative risks with two sample survival data. *Biometrika, 92,* 1–17.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics, 21,* 76–99.
Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association, 90,* 178–184.
Zhao, Y. (2005). Regression analysis for long-term survival rate via empirical likelihood. *Journal of Nonparametric Statistics, 17,* 995–1007.
Zhao, Y., Hsu, Y. S. (2005). Semiparametric analysis for additive risk model via empirical likelihood. *Communications in Statistics–Simulation and Computation, 34,* 135–143.
Zhao, Y., Qin, G. (2006). Inference for the mean residual life function via empirical likelihood. *Communications in Statistics—Theory and Methods, 35,* 1025–1036.