Ann Inst Stat Math (2011) 63:331–346
DOI 10.1007/s10463-009-0223-7

Empirical likelihood method for linear transformation models

Wen Yu · Yunting Sun · Ming Zheng

Received: 28 April 2008 / Revised: 14 October 2008 / Published online: 2 April 2009

Abstract Empirical likelihood inferential procedure is proposed for right censored survival data under linear transformation models, which include the commonly used proportional hazards model as a special case. A log-empirical likelihood ratio test statistic for the regression coefficients is developed. We show that the proposed log-empirical likelihood ratio test statistic converges to a standard chi-squared distribution. The result can be used to make inference about the entire regression coefficients vector as well as any subset of it. The method is illustrated by extensive simulation studies and a real example.

Keywords Empirical likelihood · Right censored data · Linear transformation models · Likelihood ratio test · Chi-squared distribution

1 Introduction

The proportion hazards model introduced by Cox (1972) is one of the most popular regression models used in applications and has been fully explored in theory. When the proportional hazards model is not suitable for modeling some survival data, other alternative models, such as the proportional odds model, are proposed. Both of the commonly used models are special cases of linear transformation models. Since linear transformation models include a general class of semiparametric regression

W. Yu (✉) · M. Zheng
Department of Statistics, Fudan University, 220 Handanlu, 200433 Shanghai, China
e-mail: 061025015@fudan.edu.cn

Y. Sun
Department of Statistics, Stanford University,
Sequioa Hall, 390 Serra Mall, Stanford, CA, 94305-4065, USA

models, various estimation approaches and inferential procedures were studied by many authors, cf. Cheng et al. (1995, 1997), Fine et al. (1998), etc. Among them, Chen et al. (2002) proposed an estimation method in line with Cox's partial likelihood approach to estimate the regression coefficients. The key step of their method was to construct certain martingale type estimating equations which mimicked the partial likelihood score equations for the proportional hazards model.

Empirical likelihood was proposed by Owen (1988, 1990) for constructing generalized likelihood ratio test statistics and corresponding confidence regions. It was motivated by an earlier work of Thomas and Grunkemeier (1975), which provided a way to construct confidence intervals for survival probability-related quantities through constrained likelihood ratios. Owen showed that the Wilks theorem of chi-squared limiting distribution for log-likelihood ratio statistic continues to hold for the empirical likelihood. The empirical likelihood method is well recognized to possess several desired properties including range preserving, transformation respecting, data decided shape for confidence region, implicit studentizing carried out in internal optimization without the need to estimate the variance explicitly and better small sample size performance than normal approximation method. Thus, the method has been extended to linear regression model (Owen 1991, Chen 1994) and many other areas in statistical analysis, cf. Chen and Hall (1993) on quantile estimation, Qin and Lawless (1994) on optimal linear combination of estimating equations, Chen (1996) on density estimation, among many others.

A great deal of efforts have been made to extend the empirical likelihood method to statistical analysis with censored survival data. For example, Wang and Jing (2001) discussed empirical likelihood for a class of functionals of survival distributions. Wang and Wang (2001) developed an empirical likelihood method for mean difference inference with censored data. Qin and Jing (2001b) considered empirical likelihood inferential procedure for the proportional hazards model. Zhou (2005) proposed an empirical likelihood based approach for the accelerated failure time model. Qin and Jing (2001a) and Li and Wang (2003) studied empirical likelihood for censored linear regression analysis, while Wang and Li (2002) discussed censored semiparametric regression analysis, etc. More recently, Lu and Liang (2006) developed an empirical likelihood inferential procedure for the regression coefficients in linear transformation models. They constructed a log-empirical likelihood ratio test statistic based on the martingale type estimating equations proposed by Chen et al. (2002). However, their likelihood ratio does not have a chi-squared limiting distribution. Instead, it converges to a weighted sum of independent chi-squared distributions. In order to get the quantiles of this limiting distribution, one has to estimate the unknown weights and do Monte Carlo simulations. These make the inferential procedure complicated and may introduce a cumulative bias.

In the present paper, we still consider the empirical likelihood inference for linear transformation models, based on the martingale type estimation equations proposed by Chen et al. (2002). By appropriately modifying the construction of the empirical likelihood ratio proposed by Lu and Liang (2006), we obtain a new log-empirical likelihood ratio test statistic that has a standard chi-squared limiting distribution. Based on this log-empirical likelihood ratio, the inference about the entire regression coefficients vector or any subset of it is easy to implement since there is no need to estimate variance-covariance matrices explicitly and the quantiles of standard chi-squared distribution can be used directly.

The paper is organized as follows. In Sect. 2, we set up notation and describe model specification as well as the martingale type estimating equations proposed by Chen et al. (2002). In Sect. 3, our proposed log-empirical likelihood ratio test statistic is defined and the limiting distribution is derived. Some numerical studies are presented in Sect. 4. Section 5 concludes. Technical details are given in the Appendix.

2 Notation, model specification and estimating equations

Let T be a positive continuous random variable and let Z be a p-dimensional covariate. The assumption of linear transformation models is

$$H(T) = -\beta^T Z + \varepsilon,$$

where H is an unknown monotone function, ε is a random error term with a known distribution and is independent of covariate, and β is an unknown p-dimensional regression coefficients vector of interest. T may be censored by a certain censoring variable C, which is independent of T given Z. Let $\widetilde{T} = \min(T, C)$ and $\delta = I(T \leq C)$. Let $\lambda(\cdot)$ and $\Lambda(\cdot)$ be the hazard and cumulative hazard function of ε, respectively. We have the following usual counting process notation:

$$N(t) = I(T \leq \min(t, C)) = \delta I(\widetilde{T} \leq t), \quad Y(t) = I(\widetilde{T} \geq t),$$
$$M(t) = N(t) - \int_0^t Y(s) \mathrm{d}\Lambda \left\{\beta_0^T Z + H_0(s)\right\},$$

where (β_0, H_0) are the true values of (β, H). With the assumption of the model, $M(t)$ is a martingale process under certain suitable filtration.

Let $(\widetilde{T}_i, \delta_i, Z_i)$, $i = 1, \ldots, n$ be independent observations of $(\widetilde{T}, \delta, Z)$ and $\{N_i(t), Y_i(t), M_i(t)\}$ be the sample analogues of $\{N(t), Y(t), M(t)\}$. Chen et al. (2002) proposed the following estimation equations:

$$\sum_{i=1}^{n} \int_0^{\infty} Z_i \left[\mathrm{d}N_i(t) - Y_i(t) \mathrm{d}\Lambda \left\{\beta^T Z_i + H(t)\right\}\right] = 0, \tag{1}$$

$$\sum_{i=1}^{n} \left[\mathrm{d}N_i(t) - Y_i(t) \mathrm{d}\Lambda \left\{\beta^T Z_i + H(t)\right\}\right] = 0. \tag{2}$$

They used the solution of (1) and (2) to be the estimator of (β_0, H_0), denoted by $(\widehat{\beta}, \widehat{H})$. Under some suitable regularity conditions, $\sqrt{n}(\widehat{\beta} - \beta_0)$ was shown to be asymptotically normally distributed with mean zero and variance–covariance matrix

$A(\beta_0)^{-1}V(\beta_0)(A(\beta_0)^{-1})^T$. Here

$$V(\beta_0) = \int_0^\tau E\left[(Z - \mu_Z(\beta_0, t))^{\otimes 2}\, \lambda \left\{\beta_0^T Z + H_0(t)\right\} Y(t)\right] \mathrm{d}H_0(t),$$
$$A(\beta_0) = \int_0^\tau E\left[(Z - \mu_Z(\beta_0, t))\, Z^T \lambda' \left\{\beta_0^T Z + H_0(t)\right\} Y(t)\right] \mathrm{d}H_0(t),$$

where

$$\mu_Z(\beta_0, t) = \frac{E\left[Z\lambda\left\{H_0(\widetilde{T}) + \beta_0^T Z\right\} Y(t) B(\beta_0, t, \widetilde{T})\right]}{E\left[\lambda\left\{H_0(t) + \beta_0^T Z\right\} Y(t)\right]},$$
$$B(\beta_0, t, s) = \exp\left(\int_s^t \frac{E\left[\lambda'\left\{H_0(u) + \beta_0^T Z\right\} Y(u)\right]}{E\left[\lambda\{H_0(u) + \beta_0^T Z\} Y(u)\right]} \mathrm{d}H_0(u)\right),$$

$\tau = \inf\{t : \mathrm{pr}(\widetilde{T} > t) = 0\}$, $\lambda'(t) = \mathrm{d}\lambda(t)/\mathrm{d}t$ and $b^{\otimes 2} = bb^T$ for any vector b.

3 Empirical likelihood method and asymptotic properties

Since each individual can be viewed as an observation and the data is a random sample of n independent individuals, the empirical likelihood function can be written as

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

with suitable constraints. Motivated by the estimating Eq. (1), Lu and Liang (2006) defined

$$W_{ni}(\beta) = \int_0^\infty Z_i \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left\{\beta^T Z_i + \widehat{H}(\beta, t)\right\}\right], \quad i = 1, \ldots, n,$$

where $\widehat{H}(\beta, t)$ is the solution of the estimating Eq. (2) given fixed β. In addition to the standard unit total probability constraint, they introduced the following constraint

$$\sum_{i=1}^n p_i W_{ni}(\beta) = 0$$

and consequently constructed an empirical likelihood ratio

$$R_n(\beta) = \sup\left\{\prod_{i=1}^n np_i \,\Big|\, \sum_{i=1}^n p_i W_{ni}(\beta) = 0,\ \sum_{i=1}^n p_i = 1,\ p_i \geq 0,\ i = 1, \ldots, n\right\}.$$

The asymptotic properties of this empirical likelihood ratio were explored. Define the log-empirical likelihood ratio

$$l_n(\beta) = -2 \log R_n(\beta).$$

Lu and Liang (2006) showed that under some regularity conditions, it converges to the distribution of a weighted sum of independent chi-squared random variables, that is, when $\beta_0 = \beta$, as $n \to \infty$,

$$l_n(\beta) \to l_1 \chi^2_{1,1} + \cdots + l_p \chi^2_{p,1}$$

in distribution, where $l_1, \ldots, l_p$ are the eigenvalues of $\Sigma(\beta_0)^{-1} V(\beta_0)$,

$$\Sigma(\beta_0) = \int_0^\tau E\left[Z^{\otimes 2} \lambda\{\beta_0^T Z + H_0(t)\} Y(t)\right] \mathrm{d}H_0(t),$$

and $\chi^2_{1,1}, \ldots, \chi^2_{p,1}$ are independent standard chi-squared random variables with 1 degree of freedom. In order to use this result, one has to estimate $l_1, \ldots, l_p$ first. Lu and Liang (2006) proposed to estimate them by using the eigenvalues of $\widehat{\Sigma}(\beta_0)^{-1} \widehat{V}(\beta_0)$. Here

$$\widehat{\Sigma}(\beta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau Z_i^{\otimes 2} \lambda\{\beta_0^T Z_i + \widehat{H}(\beta_0, t)\} Y_i(t) \mathrm{d}\widehat{H}(\beta_0, t),$$

$$\widehat{V}(\beta_0) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(Z_i - \overline{Z}(\beta_0, t)\right)^{\otimes 2} \lambda\{\beta_0^T Z_i + \widehat{H}(\beta_0, t)\} Y_i(t) \mathrm{d}\widehat{H}(\beta_0, t),$$

where

$$\overline{Z}(\beta_0, t) = \frac{\sum_{i=1}^n Z_i \lambda\left\{\beta_0^T Z_i + \widehat{H}(\beta_0, \widetilde{T}_i)\right\} Y_i(t) \widehat{B}(\beta_0, t, \widetilde{T}_i)}{\sum_{i=1}^n \lambda\left\{\beta_0^T Z_i + \widehat{H}(\beta_0, t)\right\} Y_i(t)}, \tag{3}$$

and

$$\widehat{B}(\beta_0, s, t) = \exp\left(\int_s^t \frac{\sum_{i=1}^n \lambda'\{\beta_0^T Z_i + \widehat{H}(\beta_0, u)\} Y_i(u)}{\sum_{i=1}^n \lambda\{\beta_0^T Z_i + \widehat{H}(\beta_0, u)\} Y_i(u)} \mathrm{d}\widehat{H}(\beta_0, u)\right).$$

After estimating the weights, one needs to do simulation to find out the quantiles of the limiting distribution.

The log-empirical likelihood ratio proposed by Lu and Liang (2006) dose not follow the standard asymptotic result for empirical likelihood method, in which the log-likelihood ratio has a standard chi-squared limiting distribution. Their explanation of such phenomenon is that because of the plugging in of $\widehat{H}(\beta, t)$, $W_{ni}(\beta)$, $i = 1, \ldots, n$ are no longer independent random variables. As we have described, when making inferences about the regression coefficients, the quantiles of the limiting distribution of their likelihood ratio need to be estimated. The procedure involves estimating some complicated variance-covariance matrices explicitly and doing Monte Carlo simulations, which may create a cumulative bias.

We find that although $W_{ni}(\beta), i = 1, \ldots, n$ are dependent, by modifying the construction of $W_{ni}(\beta)$, we are able to get a log-empirical likelihood ratio whose limiting distribution follows a standard chi-squared distribution. Specifically, define

$$\widetilde{W}_{ni}(\beta) = \int_0^\infty \left(Z_i - \overline{Z}(\beta, t)\right) \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left\{\beta^T Z_i + \widehat{H}(\beta, t)\right\}\right], \quad i = 1, \ldots, n,$$

where $\overline{Z}(\beta, t)$ is defined by (3), with β_0 replaced by β. For testing $\beta_0 = \beta$, we propose the following log-empirical likelihood ratio

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

By Language multiplier, it is easy to show that

$$\widetilde{l}_n(\beta) = 2\sum_{i-1}^n \log\left(1 + \eta(\beta)^T \widetilde{W}_{ni}(\beta)\right),$$

where $\eta(\beta)$ is the solution of the equations

$$\frac{1}{n}\sum_{i=1}^n \frac{\widetilde{W}_{ni}(\beta)}{1 + \eta^T \widetilde{W}_{ni}(\beta)} = 0. \tag{4}$$

The main reason of constructing $\widetilde{W}_{ni}(\beta)$ is motivated by the so-called i.i.d. representation of the estimating equations. In the Appendix of Chen et al. (2002), they showed that under some regularity conditions,

$$\sum_{i=1}^n \int_0^\infty Z_i \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left\{\beta_0^T Z_i + \widehat{H}(\beta_0, t)\right\}\right] = \sum_{i=1}^n \widetilde{W}_i(\beta_0) + o_p(n^{1/2}),$$

where

$$\widetilde{W}_i(\beta_0) = \int_0^\tau (Z_i - \mu_Z(\beta_0, t)) \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left\{\beta_0^T Z_i + H_0(t)\right\}\right], \quad i=1, \ldots, n$$

are i.i.d. random vectors. Thus, for each i and fixed β, $\widetilde{W}_{ni}(\beta)$ is the corresponding quantity to $\widetilde{W}_i(\beta)$ with the unknown parameters replaced by their consistent estimators. We find that by imposing constraint in terms of $\widetilde{W}_{ni}(\beta)$, the proposed log-empirical likelihood ratio test statistic has a standard chi-squared limiting distribution under the null hypothesis. To see this, we need the following results.

Proposition 1 *Under suitable regularity conditions, we have that*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0) \rightarrow N(0, V(\beta_0)),$$

in distribution, as $n \rightarrow \infty$*. Moreover, we have that*

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T \rightarrow V(\beta_0),$$

in probability as $n \rightarrow \infty$*.*

By the results of Proposition 1, the limiting distribution of $\widetilde{l}_n(\beta)$ is shown by the following proposition.

Proposition 2 *Assume that the similar regularity conditions given in Proposition 1 hold. Under the null hypothesis* $\beta_0 = \beta$*,*

$$\widetilde{l}_n(\beta) \rightarrow \chi_p^2$$

in distribution, as $n \rightarrow \infty$*, where* χ_p^2 *is a standard chi-squared random variable with* p *degree of freedom.*

Then Proposition 2 enables us to make global inference about β_0 by the standard chi-squared distribution. For $0 < \alpha < 1$, define $I_\alpha = \{\beta,\ \widetilde{l}_n(\beta) \leq \chi_p^{2,1-\alpha}\}$, where $\chi_p^{2,1-\alpha}$ is the $(1-\alpha)$th quantile of χ_p^2. Then I_α is the asymptotic $(1-\alpha)100\%$ confidence region for β_0.

Sometimes we may be interested in certain subvector of the entire parameter vector. Write $\beta_0 = (\beta_{10}^T, \beta_{20}^T)^T$ and suppose that we are only interested in the q-dimensional subvector β_{10}. In order to make inference about β_{10}, we may turn to the profile empirical likelihood method for handling nuisance parameters. Specifically, define

$$\widetilde{l}_n^*(\beta_1) = \inf_{\beta_2}\widetilde{l}_n(\beta_1, \beta_2).$$

The function $\widetilde{l}_n^*(\beta_1)$ is usually called the profile log-empirical likelihood ratio. Note that such profile likelihoods are widely used in parametric models with nuisance parameters and the profile parametric likelihood ratio has a chi-squared limiting distribution. The similar result still holds for the proposed empirical likelihood ratio here. The following proposition gives out formal presentation.

Proposition 3 *Assume that the similar regularity conditions given in Proposition 1 hold. Under the null hypothesis* $\beta_{10} = \beta_1$*, as* $n \rightarrow \infty$*,*

$$\widetilde{l}_n^*(\beta_1) \rightarrow \chi_q^2$$

in distribution, where χ_q^2 *is a standard chi-squared random variable with* q *degree of freedom.*

Proposition 3 enables us to construct confidence region for β_{10}. For $0 < \alpha < 1$, define $I_\alpha^* = \{\beta_1, \widetilde{l}_n^*(\beta_1) \le \chi_q^{2,1-\alpha}\}$, where $\chi_q^{2,1-\alpha}$ is the $(1-\alpha)$th quantile of χ_q^2. Then I_α^* is the asymptotic $(1-\alpha)100\%$ confidence region for β_{10}.

Making inference about subsets of the entire regression coefficients vector is easier to complement by applying Proposition 3 than using the corresponding method proposed by Lu and Liang (2006). In their work, when making inference about the subvector, one has to redefine the log-likelihood ratio by changing the construction of $W_{ni}(\beta)$ and the limiting distribution of that log-likelihood ratio test statistic is still a weighted sum of standard chi-squared distributions with the unknown weights to be estimated.

4 Numerical studies

The properties and behaviors of finite sample size of our proposed empirical likelihood inference were assessed in a series of simulation studies. We chose $H(t) = \log(t)$ and let the hazard function of ε be of the form $\lambda(t) = \exp(t)/\{1 + r\exp(t)\}$ with $r = 0, 0.5, 1, 1.5, 2$ (Dabrowska and Doksum 1988). Note that the proportional hazards and proportional odds models correspond to $r = 0$ and $r = 1$, respectively. For each model, two independent covariates Z_1 and Z_2 were generated with Z_1 following the Bernoulli distribution with success probability 0.5 and Z_2 the uniform distribution on (0, 1). We set $\beta_0 = (0, 1)^T$. Two different censoring schemes were considered. One was covariate independent censoring with censoring variable generated from the uniform distribution on $(0, c)$. The other was covariate dependent censoring with censoring variable set to be $-Z_1 - Z_2 + \text{Un}(0, c)$, where $\text{Un}(0, c)$ represents the uniform distribution on $(0, c)$. Note that the proposed method is valid for both covariate independent and dependent censoring. In both cases, the constant c was chosen to yield certain censoring proportion. The sample size was 100 and all the simulations were based on 1,000 replications.

The first set of simulation studies was conducted for global inference. We tested the null hypothesis $\beta_0 = (0, 1)^T$. The rejection region based on our proposed log-empirical likelihood ratio could be constructed as $R_1 = \{\widetilde{l}_n(\beta) > \chi_2^{2,1-\alpha}\}$, where $\widetilde{l}_n(\beta)$ and $\chi_2^{2,1-\alpha}$ are defined in Sect. 3. The nominal levels of the type I error were chosen to be 0.05 and 0.1. Besides the proposed log-empirical likelihood ratio test statistic (ELR), we also simulated the empirical type I errors of the Wald-type test statistic based on the normal approximation method proposed by Chen et al. (2002) (NA) and those of the log-empirical likelihood ratio test statistic proposed by Lu and Liang (2006) (LL) for comparison.

The second set of simulation studies was conducted for testing subvector of the entire regression coefficients vector. Write $\beta_0 = (\beta_{10}, \beta_{20})^T$, and then $\beta_{10} = 0$, $\beta_{20} = 1$. The null hypothesis we hoped to test was $\beta_{20} = 1$. The rejection region based on our proposed log-empirical likelihood ratio could be written as $R_2 = \{\widetilde{l}_n^*(\beta_2) > \chi_1^{2,1-\alpha}\}$, where $\widetilde{l}_n^*(\beta_2)$ and $\chi_1^{2,1-\alpha}$ are defined in Sect. 3. The nominal levels of type I error were chosen to be 0.05 and 0.1. The Wald-type test statistic and Lu and Liang (2006)'s log-empirical likelihood ratio test statistic for covariate adjustment were also calculated.

Tables 1 and 2 summarize the results from the first and second simulation studies, respectively. From Table 1, we find that all the three tests give the appropriate empirical type I errors in nearly all cases. The presence of censoring and the scheme of censoring do not affect the performances of the test statistics. In general, the normal approximation based test is more conservative than the two empirical likelihood based tests. When $r = 0$ and 0.5, the empirical type I errors of the normal approximation based test are usually a little bit closer to the corresponding nominal levels than those of the empirical likelihood based tests. When r becomes larger, the empirical likelihood based tests seem to be more accurate. The empirical type I errors of the proposed empirical likelihood ratio test are comparable with those of Lu and Liang (2006)'s test. Moreover, as mentioned above, our proposed inferential procedure is easier to implement. From Table 2, we obtain similar observations. The proposed empirical likelihood ratio test provides satisfactory empirical type I errors in all cases.

We then applied the proposed method to a study on multiple myeloma reported by Krall et al. (1975). In the study, 65 patients were treated with alkylating agents. 48 of them died during the study and 17 survived. This data is regarded as a main example in the PROC PHREG of the SAS/STAT User's Guide (SAS Institute, Inc. 1999, pp. 2608–2617, 2536–2641) to focus on the proportional hazards model with the logarithm of blood urea nitrogen, LogBUN and haemoglobin, HGB, as the covariates. Here

Table 1 Empirical type I errors for testing $\beta_0 = (0, 1)^T$ with the nominal levels chosen to be 0.05 and 0.1

Censoring scheme	Censoring proportion (%)	Method	$r = 0$		$r = 0.5$		$r = 1$		$r = 1.5$		$r = 2$	
			0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Covariate	10	NA	0.041	0.089	0.038	0.078	0.033	0.080	0.034	0.065	0.032	0.072
independent		LL	0.064	0.128	0.056	0.103	0.056	0.125	0.061	0.118	0.056	0.099
		ELR	0.059	0.115	0.063	0.103	0.060	0.098	0.057	0.109	0.051	0.103
	30	NA	0.042	0.092	0.036	0.097	0.036	0.093	0.031	0.068	0.039	0.081
		LL	0.061	0.118	0.069	0.119	0.062	0.114	0.049	0.102	0.050	0.103
		ELR	0.061	0.113	0.056	0.114	0.055	0.107	0.055	0.098	0.053	0.104
	50	NA	0.051	0.090	0.043	0.081	0.036	0.090	0.030	0.079	0.039	0.074
		LL	0.050	0.098	0.053	0.113	0.061	0.116	0.046	0.117	0.053	0.117
		ELR	0.058	0.111	0.042	0.090	0.057	0.117	0.054	0.109	0.044	0.098
Covariate	10	NA	0.042	0.090	0.046	0.107	0.032	0.071	0.038	0.079	0.037	0.084
dependent		LL	0.062	0.133	0.065	0.129	0.063	0.123	0.055	0.109	0.062	0.125
		ELR	0.067	0.122	0.064	0.118	0.061	0.105	0.063	0.099	0.056	0.112
	30	NA	0.043	0.095	0.047	0.101	0.031	0.071	0.036	0.077	0.037	0.087
		LL	0.067	0.128	0.064	0.111	0.047	0.096	0.053	0.105	0.052	0.105
		ELR	0.068	0.118	0.057	0.109	0.052	0.096	0.051	0.102	0.041	0.104
	50	NA	0.050	0.108	0.051	0.098	0.028	0.071	0.036	0.091	0.039	0.086
		LL	0.073	0.114	0.060	0.117	0.061	0.113	0.058	0.116	0.052	0.109
		ELR	0.066	0.120	0.048	0.103	0.052	0.099	0.050	0.095	0.048	0.094

Table 2 Empirical type I errors for testing $\beta_{20} = 1$ with the nominal levels chosen to be 0.05 and 0.1

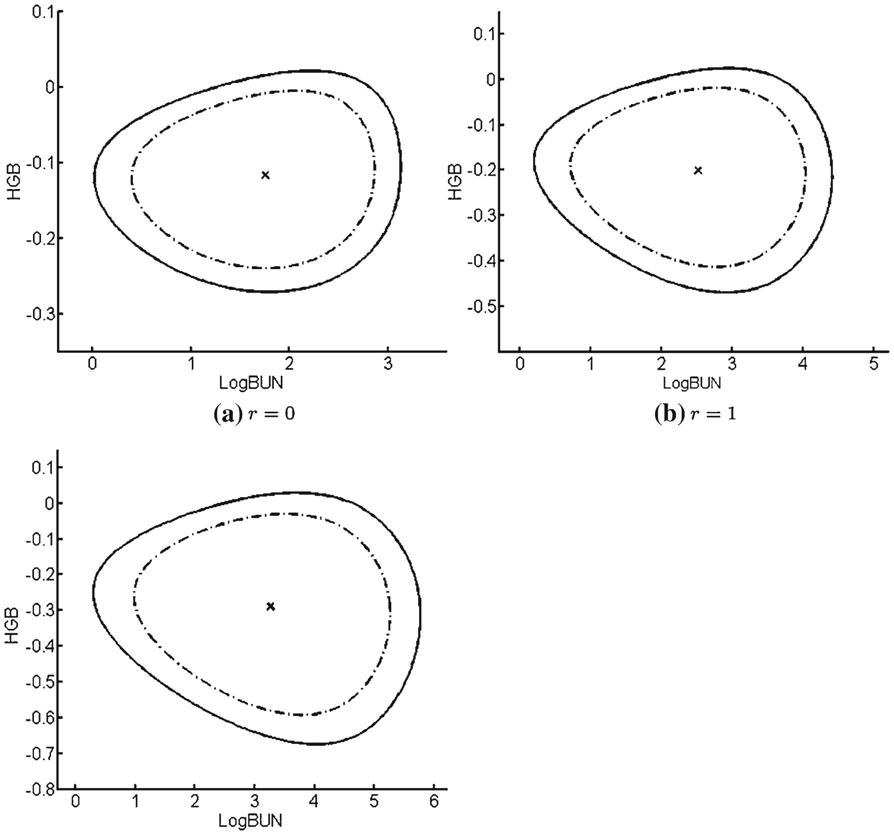
Censoring scheme	Censoring proportion (%)	Method	$r = 0$		$r = 0.5$		$r = 1$		$r = 1.5$		$r = 2$	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	
Covariate independent	10	NA	0.059	0.099	0.045	0.096	0.041	0.085	0.035	0.085	0.034	0.076
		LL	0.068	0.119	0.057	0.121	0.062	0.123	0.061	0.116	0.063	0.108
		ELR	0.063	0.109	0.059	0.110	0.059	0.112	0.051	0.115	0.051	0.106
	30	NA	0.053	0.095	0.048	0.099	0.042	0.094	0.037	0.084	0.031	0.079
		LL	0.059	0.111	0.063	0.112	0.054	0.122	0.058	0.118	0.063	0.108
		ELR	0.057	0.108	0.058	0.110	0.055	0.116	0.053	0.112	0.056	0.104
	50	NA	0.054	0.090	0.037	0.082	0.045	0.088	0.040	0.085	0.036	0.089
		LL	0.061	0.102	0.055	0.097	0.054	0.098	0.056	0.109	0.056	0.109
		ELR	0.062	0.096	0.053	0.095	0.051	0.089	0.053	0.103	0.053	0.106
Covariate dependent	10	NA	0.059	0.100	0.045	0.096	0.038	0.081	0.036	0.087	0.039	0.077
		LL	0.070	0.120	0.052	0.115	0.053	0.109	0.063	0.122	0.062	0.118
		ELR	0.068	0.119	0.052	0.102	0.055	0.108	0.056	0.113	0.053	0.110
	30	NA	0.051	0.095	0.046	0.091	0.038	0.078	0.032	0.073	0.033	0.079
		LL	0.068	0.113	0.051	0.099	0.058	0.114	0.051	0.106	0.057	0.101
		ELR	0.063	0.105	0.053	0.098	0.058	0.106	0.054	0.101	0.061	0.103
	50	NA	0.048	0.112	0.043	0.088	0.031	0.085	0.037	0.072	0.038	0.079
		LL	0.065	0.118	0.053	0.107	0.052	0.111	0.056	0.100	0.053	0.101
		ELR	0.057	0.117	0.049	0.103	0.055	0.108	0.058	0.099	0.049	0.094

we still used these two covariates and made inferences about the two corresponding regression coefficients.

As in the simulation studies, the hazard function of ε was of the form $\lambda(t) = \exp(t)/\{1 + r\exp(t)\}$ and we chose $r = 0, 1, 2$. Figure 1 shows the joint confidence regions for the two regression coefficients, which were obtained based on the Proposition 2 in Sect. 3. The outer loops are the 95% confidence regions. The centers of the confidence regions, at which the empirical likelihood ratio attains 1, are the Chen et al. (2002)'s estimators for the regression coefficients. They are marked by X on the plots. We see that different from the normal approximation based confidence regions, the empirical likelihood based ones do not have elliptical shapes. The shapes are decided by the data themselves.

The empirical likelihood based confidence intervals for each coefficient could be obtained as the left (right, upper or lower) most points of the contours decided by quantiles of χ_1^2 distribution (the inner loops correspond to 0.95 quantile). We also fitted the proportional hazards model and compared the normal approximated confidence intervals based on partial likelihood with the ones based on our proposed empirical likelihood ratio. The results are listed in Table 3.

We find that when $r = 0$, that is, the corresponding transformation model is just the proportional hazards model, the empirical likelihood ratio based confidence intervals are numerically close to the normal approximated ones. The intervals become longer when $r = 1$ and 2, while the inference conclusions for each coefficient do not change.



(a) $r = 0$ (b) $r = 1$ (c) $r = 2$

Fig. 1 Empirical likelihood based confidence regions for multiple myeloma data

Table 3 Confidence intervals for multiple myeloma data

		NA	$r = 0$	$r = 1$	$r = 2$
90%	LogBUN	(0.7260, 2.7870)	(0.6156, 2.7069)	(1.0223, 3.7963)	(1.3854, 4.9392)
	HGB	(−0.2113, −0.0227)	(−0.2199, −0.0224)	(−0.3778, −0.0460)	(−0.5407, −0.0681)
95%	LogBUN	(0.5292, 2.9788)	(0.3950, 2.8669)	(0.7035, 4.0346)	(0.9733, 5.2614)
	HGB	(−0.2293, −0.0047)	(−0.2401, −0.0049)	(−0.4143, −0.0176)	(−0.5929, −0.0291)

5 Concluding remarks

Similar to Lu and Liang (2006), the proposed empirical likelihood method for linear transformation models generalizes the test of Owen (1991) for linear models. Compared with their work, our method preserves the important feature of the standard empirical likelihood method, that is, the log-empirical likelihood ratio has a standard chi-squared limiting distribution. Thus, when making inference about the regression coefficients, there is no need to estimate complicated variance-covariance matrices explicitly. Moreover, the profile empirical likelihood can be applied to make inference about any subset of the entire regression coefficients vector.

However, since the estimating equations proposed by Chen et al. (2002) is not efficient except for the proportional hazards model, correspondingly the proposed empirical likelihood test is not efficient. The closed form of the relative efficiency of our proposed test to that of Lu and Liang (2006) is unknown.

Our method may be generalized to the accelerated failure time model and other regression models for survival data. These will be future research topics.

Acknowledgments The authors would like to thank the referees for their helpful comments and suggestions which greatly improved the original version of the paper. The research is supported by Shanghai Leading Academic Discipline Project, project number: B210.

Appendix

In the appendix, we prove Proposition 1–3 presented in Sect. 3. Regularity conditions for ensuring the central limit theorem for counting process martingales such as those assumed in Fleming and Harrington (1991) are assumed here without specific statement. Similar to Chen et al. (2002), we assume that τ is finite, $P(T > \tau) > 0$ and $P(C \geq \tau) = P(C = \tau) > 0$, which is to avoid a lengthy technical discussion about the tail behavior. Let $\psi(t) = \lambda'(t)/\lambda(t)$. We assume the positivity of $\lambda(\cdot)$, the continuity of $\psi(\cdot)$ and that $\lim_{s\to-\infty} \lambda(s) = 0 = \lim_{s\to-\infty} \psi(s)$. We assume that $P(\|Z\| < C) = 1$ for some constant $C > 0$ and that $H_0(\cdot)$ has continuous and positive derivatives. Furthermore, $A(\beta_0)$ and $V(\beta_0)$ are assumed to be finite and nondegenerate.

A.1. Proof of Proposition 1

Recall that $\widehat{H}(\beta_0, t)$ satisfies

$$\sum_{i=1}^{n} \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\{\beta_0^T Z_i + \widehat{H}(\beta_0, t)\} \right] = 0,$$

it is easy to show that

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{W}_{ni}(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0) \\
&\quad + \int_0^{\infty} \overline{Z}(\beta_0, t) \sum_{i=1}^{n} \left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left(\beta_0^T Z_i + \widehat{H}(\beta_0, t)\right) \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(\beta_0).
\end{aligned}$$

Thus, by the proof of Proposition 2 in Lu and Liang (2006), we get that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0) \rightarrow N(0, V(\beta_0))$$

in distribution as $n \rightarrow \infty$.

It can be shown that for each i,

$$\begin{aligned}\widetilde{W}_{ni}(\beta_0) &= \widetilde{W}_i(\beta_0) - \int_0^{\tau}\left(\overline{Z}(\beta_0, t) - \mu_Z(t)\right)\left[\mathrm{d}N_i(t) - Y_i(t)\mathrm{d}\Lambda\left(\beta_0^T Z_i + H_0(t)\right)\right] \\ &\quad - \int_0^{\tau}\left(Z_i - \overline{Z}(\beta_0, t)\right) Y_i(t)\mathrm{d}\left[\Lambda\left(\beta_0^T Z_i + \widehat{H}(\beta_0, t)\right) - \Lambda\left(\beta_0^T Z_i + H_0(t)\right)\right] \\ &=: \widetilde{W}_i(\beta_0) + r_{i1} + r_{i2}.\end{aligned}$$

By the similar arguments in the appendix of Chen et al. (2002), for each $i = 1, \ldots, n$, we have

$$\|r_{i1}\| \le C_1 \sup_{0\le t\le\tau}\left\|\overline{Z}(\beta_0, t) - \mu_Z(t)\right\| = o_p(1),$$

and

$$\begin{aligned}\|r_{i2}\| &\le C_2\left\|\frac{\lambda\left\{H_0(\widetilde{T}_i) + \beta_0^T Z_i\right\}}{\lambda^*\{H_0(\widetilde{T}_i)\}}\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^{\tau} Y_i(t)\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\mathrm{d}M_j(t) + o_p(n^{-1/2})\right)\right\| \\ &= o_p(1),\end{aligned}$$

where C_1 and C_2 are positive constants. For any $a \in \mathbf{R}^p$, we have the following decompositions:

$$\begin{aligned}&a^T\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T - \frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_i(\beta_0)\widetilde{W}_i(\beta_0)^T\right)a \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[a^T\left(\widetilde{W}_{ni}(\beta_0) - \widetilde{W}_i(\beta_0)\right)\right]^2 + \frac{2}{n}\sum_{i=1}^{n}\left(a^T\widetilde{W}_i(\beta_0)\right) \\ &\quad\times\left[a^T\left(\widetilde{W}_{ni}(\beta_0) - \widetilde{W}_i(\beta_0)\right)\right].\end{aligned}$$

From the order of r_{i1} and r_{i2} derived above, it can be shown that the two parts on the right-hand side of the above equation are both $o_p(1)$. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T = \frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_i(\beta_0)\widetilde{W}_i(\beta_0)^T + o_p(1).$$

By the law of large number,

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_i(\beta_0)\widetilde{W}_i(\beta_0)^T \to E\left(\widetilde{W}(\beta_0)\widetilde{W}(\beta_0)^T\right) = V(\beta_0)$$

in probability as $n \to \infty$. Therefore

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T \to V(\beta_0)$$

in probability as $n \to \infty$.

A.2. Proof of Proposition 2

Similar to Lu and Liang (2006), we need to show that (i) $\max_{1\le i\le n}\|\widetilde{W}_{ni}(\beta_0)\| = o_p(n^{1/2})$ and (ii) $\eta = O_p(n^{-1/2})$, where η is the solution of (4). It has been shown that for each i

$$\widetilde{W}_{ni}(\beta_0) = \widetilde{W}_i(\beta_0) + r_{i1} + r_{i2}.$$

Since $\widetilde{W}_i(\beta_0)$, $i = 1, \ldots, n$ are iid random variables with finite second moment, by the lemma 11.2 of Owen (2001), $\max_{1\le i\le n}\|\widetilde{W}_i(\beta_0)\| = o_p(n^{1/2})$. By the above arguments, r_{i1} and r_{i2} are both of order $o_p(1)$. Thus (i) follows. Moreover, by applying similar arguments in the proof of the Theorem 3.2 of Owen (2001), we can derive that (ii) is valid.

Combining (i), (ii) and arguments in the same reference mentioned above, it can be shown that

$$\eta = \left(\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T\right)^{-1}\left(\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\right) + o_p(n^{-1/2}).$$

Thus, by Taylor expansion, we have

$$\begin{aligned}\widetilde{l}_n(\beta_0) &= 2\sum_{i=1}^{n}\eta^T\widetilde{W}_{ni}(\beta_0) - \sum_{i=1}^{n}\eta^T\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T\eta + o_p(1)\\ &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\right)^T\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\widetilde{W}_{ni}(\beta_0)^T\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{W}_{ni}(\beta_0)\right)\\ &\quad + o_p(1).\end{aligned}$$

By the results of Proposition 1 and the Slutsky lemma, Proposition 2 follows.

A.3. Proof of Proposition 3

Corresponding to $(\beta_{10}^T, \beta_{20}^T)^T$, write $Z = (Z_1^T, Z_2^T)^T$. Define

$$\widetilde{A}(\beta_0) = \int_0^\tau E\left[\{Z - \mu_Z(\beta_0, t)\}Z_2^T \lambda'\{\beta_0^T Z + H_0(t)\}Y(t)\right] \mathrm{d}H_0(t),$$

Since $A(\beta_0)$ is assumed to be nondegenerate, $\widetilde{A}(\beta_0)$ is of rank $p - q$. Let $\widehat{\beta}_2 = \arg\inf_{\beta_2} \widetilde{l}_n(\beta_{10}, \beta_2)$. By the similar arguments in Qin and Lawless (1994) and Chen et al. (2002), we can show that

$$\sqrt{n}(\widehat{\beta}_2 - \beta_{20}) = -W^{-1}\widetilde{A}(\beta_0)^T V(\beta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0) + o_p(1),$$

$$\sqrt{n}\eta_2 = \left(I - V(\beta_0)^{-1}\widetilde{A}(\beta_0)W^{-1}\widetilde{A}(\beta_0)^T\right) V(\beta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0) + o_p(1),$$

where η_2 is the corresponding Lagrange multiplier and

$$W = \widetilde{A}(\beta_0)^T V(\beta_0)^{-1}\widetilde{A}(\beta_0).$$

Thus, by Taylor expansion,

$$\begin{aligned}\widetilde{l}_n^*(\beta_{10}) &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0)\right)^T \left(V(\beta_0)^{-1} - V(\beta_0)^{-1}\widetilde{A}(\beta_0)W^{-1}\widetilde{A}(\beta_0)^T V(\beta_0)^{-1}\right) \\ &\quad \times \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0)\right) + o_p(1) \\ &= \left(V(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0)\right)^T S\left(V(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0)\right) + o_p(1),\end{aligned}$$

where

$$S = I - V(\beta_0)^{-1/2}\widetilde{A}(\beta_0)W^{-1}\widetilde{A}(\beta_0)^T V(\beta_0)^{-1/2}$$

is a symmetric and idempotent matrix with trace q. By Proposition 1,

$$V(\beta_0)^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{W}_{ni}(\beta_0) \to N(0, I)$$

in distribution as $n \to \infty$, Proposition 3 follows.

References

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

Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, *83*, 329–341.

Chen, S. X., Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, *21*, 621–637.

Chen, K., Jin, Z., Ying, Z. (2002). Semiparametric of transformation models with censored data. *Biometrika*, *89*, 659–668.

Cheng, S. C., Wei, L. J., Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, *82*, 835–845.

Cheng, S. C., Wei, L. J., Ying, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *Journal of the American Statistical Association*, *92*, 227–235.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society B*, *34*, 187–220.

Dabrowska, D. M., Doksum, K. A. (1988). Estimation and testing in the two-sample generalized odds rate model. *Jounral of the American Statistical Association*, *83*, 744–749.

Fine, J., Ying, Z., Wei, L. J. (1998). On the linear transformation model for censored data. *Biometrika*, *85*, 980–986.

Fleming, T. R., Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Krall, J. M., Uthoff, V. A., Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrika*, *31*, 49–57.

Li, G., Wang, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statistica Sinica*, *13*, 51–68.

Lu, W., Liang, Y. (2006). Empirical likelihood inference for linear transformation models. *Journal of Multivariate Analysis*, *97*, 1586–1599.

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

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

Owen, A. B. (1991). Empirical likelihood for linear model. *The Annals of Statistics*, *19*, 1725–1747.

Owen, A. B. (2001). *Empirical likelihood*. London: Chapman and Hall.

Qin, G., Jing, B.-Y. (2001a). Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics*, *28*, 661–673.

Qin, G., Jing, B.-Y. (2001b). Empirical likelihood for cox regression model under random censorship. *Communications in Statistics-Simulation and Computation*, *30*, 79–90.

Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, *22*, 300–325.

SAS Institute, Inc. (1999). *SAS/STAT User's Guide*, Version 8. Cary: SAS Institute Inc.

Thomas, D. R., Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, *70*, 865–871.

Wang, Q.-H., Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distributions with censored data. *Annals of the Institute of Statistical Mathematics*, *53*(3), 517–527.

Wang, Q.-H., Li, G. (2002). Empirical likelihood semiparametric regression analysis under random censorship. *Journal of Multivariate Analysis*, *83*, 469–486.

Wang, Q.-H., Wang, J.-L. (2001). Inference for the mean difference in the two-sample random censorship model. *Journal of Multivariate Analysis*, *79*(2), 295–315.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, *92*, 492–498.