Communications in Statistics—Simulation and Computation®, 36: 593–605, 2007

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610910701238301

Taylor & Francis
Taylor & Francis Group

Reliability Analysis

Test-Based Interval Estimation Under the Accelerated Failure Time Model

YICHUAN ZHAO[1] AND YIJIAN HUANG[2]

[1]Department of Mathematics and Statistics, Georgia State University, Atlanta, Georgia, USA
[2]Department of Biostatistics, School of Public Health, Emory University, Atlanta, Georgia, USA

The accelerated failure time (AFT) model is an important regression tool to study the association between survival time and covariates. Semiparametric inference procedures have been proposed extensively in the literature. Recently, Zhou (2005a) proposed to use a model-based empirical likelihood approach to interval estimation for the AFT model. However, comparison was not made with more standard approaches based on score test and Wald-type test. In this article, we conduct extensive simulation studies to evaluate their relative performance in small samples. In addition to these methods, we also devise and consider model-free empirical likelihood approach for the comparisons. Our simulation results suggest that these empirical likelihood-based methods are not better than the standard approach based on the score test.

Keywords Confidence region; Counting process; Estimating equation; Right-censoring; Wilks's theorem.

Mathematics Subject Classification Primary 62G05; Secondary 62F12.

1. Introduction

In survival analysis, the accelerated failure time (AFT) model is an important alternative to the popular proportional hazards model of Cox (1972). The AFT model relates the logarithm of the failure time linearly to the covariates, with the model error distribution unspecified. One standard inference procedure is based on the class of weighted log-rank estimating functions (Tsiatis, 1990). As shown by Ying (1993), among others, there exists a zero-crossing to a weighted log-rank estimating function that is consistent and asymptotically normal. Note that Ritov (1990) extended the estimating function of Buckley and James (1979) and showed his class is asymptotically equivalent to that of Tsiatis (1990).

Received April 7, 2006; Accepted December 8, 2006
Address correspondence to Yichuan Zhao, Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA; E-mail: yzhao@mathstat.gsu.edu

However, weighted log-rank estimating functions are neither monotone in general nor continuous, which gives rise to difficulty with root-finding and variance estimation. As an exception, the Gehan estimating function is monotone (Fygenson and Ritov, 1994) and its root-finding can be carried out with the linear programming technique (Lin et al., 1998). Recently, Jin et al. (2003) developed a broad class of monotone estimating functions and the resulting estimators represent the consistent roots of the weighted log-rank estimating functions. The difficulty with variance estimation is due to the fact that the weighted log-rank estimating functions are not differentiable. Therefore, the sandwich variance estimation does not apply. To address this issue, Parzen et al. (1994) developed a resampling approach and Huang (2002) proposed a computationally more efficient sample-based method. Nevertheless, standard approaches to confidence region by inverting the score and Wald-type tests still apply to the weighted log-rank estimating functions.

Empirical likelihood (EL) method is a powerful nonparametric method. In general, EL has unique features, such as range respecting, transformation-preserving, asymmetric confidence interval, and Bartlett correctability (Owen, 2001). The EL approach does not require to estimate the limiting covariance matrices. Moreover, the confidence region is adapted to the data set and not necessarily symmetric. Thus, it reflects the nature of the underlying data and hence gives a more representative way to make inferences about the parameter of interest. In analysis of censored survival times, for example, empirical likelihood was used to derived pointwise confidence intervals for survival function with right-censored data as early as 1975 (Thomas and Grunkemeier, 1975).

In linear regression analysis via EL for right-censored survival data, recent work includes Qin and Jing (2001) and Li and Wang (2003), among others. More recently, Zhou and Li (2004) and Zhou (2005a) developed model-based empirical likelihood confidence regions for regression parameter based on Buckley–James and rank-based estimating equations. They do not need to estimate the variance matrix when building confidence region. However, the constrained maximization of the EL has no closed form and involves n nonlinear equations, where n is a sample size. Since the computation is demanding, they applied the modified EM algorithm (cf. Zhou, 2005b) to obtain it. Moreover, the comparison was not made with standard approached based on score test and Wald-type test.

In this article, we build alternative confidence regions for regression parameter and conduct a comparison study. The rest of the article is organized as follows. In Sec. 2, we first introduce standard confidence regions for regression parameter using score test and Wald-type test. Then, we construct model-free empirical likelihood confidence regions for the regression parameter. The corresponding constrained maximization of the empirical likelihood can be done reliably by Newton–Raphson method. In Sec. 3, we conduct extensive simulation studies to compare the relative performance with other methods. The proof is presented in the Appendix.

2. Confidence Regions

2.1. *Standard Confidence Regions*

Let $T_1, \ldots, T_n$ be a sequence of positive random variables, usually representing survival (failure) times of n patients (items) in a medical study. Let $Z_1, \ldots, Z_n$ be their corresponding $(p \times 1)$ covariates sequence. The AFT model is to relate the logarithms of survival times, $\log T_i$, to their covariates through a system of linear regression equations.

$$\log T_i = \beta^{\mathrm{T}} Z_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where β is a $p \times 1$ parameter vector and ϵ_i are independent error terms with a common, but completely unspecified, distribution. There exist censoring times C_i, such that we can only observe $X_i = T_i \wedge C_i$, $\delta_i = I(T_i \leq C_i)$ and Z_i, $i = 1, \ldots, n$.

Considering the model (1), the following conditional independent censoring is assumed: Conditional on Z_i, C_i is independent of T_i, $i = 1, \ldots, n$. We define $e_i(\beta) = \log X_i - \beta^{\mathrm{T}} Z_i$, $N_i(\beta; t) = \delta_i I(e_i(\beta) \leq t)$, and $R_i(\beta; t) = I(e_i(\beta) \geq t)$. Write

$$S^{(0)}(\beta; t) = n^{-1} \sum_{i=1}^{n} R_i(\beta; t),$$

$$S^{(1)}(\beta; t) = n^{-1} \sum_{i=1}^{n} R_i(\beta; t) Z_i.$$

The weighted log-rank estimating function for β is

$$U(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta; t)(Z_i - \overline{Z}(\beta; t) dN_i(\beta; t), \tag{2}$$

where $\overline{Z}(\beta; t) = S^{(1)}(\beta, t)/S^{(0)}(\beta; t)$, and ϕ is a possibly data-dependent weight function satisfying Condition 5 of Ying (1993, p. 90). The choice of $\phi = 1$ and $\phi = S^{(0)}$ corresponds to the log-rank (Mantel, 1966) and Gehan (1965) statistics, respectively.

Write $\hat{\beta}_\phi$ and β_0 as the estimated and true values of β, respectively. Then as shown in Ying (1993), under the regularity conditions, the random vector

$$n^{1/2}(\hat{\beta}_\phi - \beta_0) \xrightarrow{\mathscr{D}} N(0, D_\phi^{-1} V_\phi D_\phi^{-1}),$$

where $D_\phi = \lim_{n \to \infty} \widehat{D}_\phi(\hat{\beta}_\phi)$ and $V_\phi = \lim_{n \to \infty} \widehat{V}_\phi(\hat{\beta}_\phi)$; here,

$$\widehat{D}_\phi(\hat{\beta}_\phi) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\hat{\beta}_\phi; t)\{Z_i - \overline{Z}(\hat{\beta}_\phi; t)\}^{\otimes 2}\{\lambda'(t)/\lambda(t)\} dN_i(\hat{\beta}_\phi; t),$$

$$\widehat{V}_\phi(\hat{\beta}_\phi) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi^2(\hat{\beta}_\phi; t)\{Z_i - \overline{Z}(\hat{\beta}_\phi; t)\}^{\otimes 2} dN_i(\hat{\beta}_\phi; t),$$

respectively, where $\lambda(\cdot)$ is the common hazard function of the error terms. Thus using the test-based approach (cf. Wei et al., 1990), an asymptotic $100(1-\alpha)\%$ Wald-type confidence region for β is given by

$$\mathscr{R}_1 = \{\beta : n^{-1} U(\beta)^{\mathrm{T}} \widehat{V}_\phi^{-1}(\hat{\beta}_\phi) U(\beta) \leq \chi_p^2(\alpha)\}, \tag{3}$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of the chi-square distribution with degrees of freedom p. Apparently the coverage accuracy of $\mathscr{R}_1$ mainly depends on the large-sample normal approximation, which might be affected by sample size and censoring rate.

In (3) for the $\widehat{V}_{\phi}(\hat{\beta}_{\phi})$, we substitute $\hat{\beta}_{\phi}$ with β. Thus, an asymptotic $100(1-\alpha)\%$ score test based confidence region for β is given by

$$\mathscr{R}_2 = \{\beta : n^{-1}U(\beta)^{\mathrm{T}}\widehat{V}_{\phi}^{-1}(\beta)U(\beta) \leq \chi_p^2(\alpha)\}, \tag{4}$$

where $\chi_p^2(\alpha)$ is defined as before.

2.2. *Model-Free EL Confidence Region*

Zhou (2005a) proposed a model-based empirical likelihood testing procedure for the rank-based estimator where the likelihood is defined as the censored empirical likelihood of the error variables $e_i(\beta)$, $i = 1, \ldots, n$, used by Thomas and Grunkemeier (1975) and Li (1995), among others. He showed the limiting distribution of the log empirical likelihood ratio at the true value of regression parameter is a chi-squared distribution. Based on the result, he builded confidence region for the regression parameter.

Now consider an alternative approach based on EL. Let $\widehat{\Lambda}(t)$ be the Nelson–Aalen estimator of cumulative hazard function $\Lambda(t) = \int_{-\infty}^{t} \lambda(s)ds$ of ϵ_i. Denote

$$\widehat{M}_i(\beta_0; t) = N_i(\beta_0; t) - \int_{-\infty}^{t} R_i(\beta_0; s)d\widehat{\Lambda}(s). \tag{5}$$

By simple algebra, we have

$$U(\beta_0) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta_0; t)(Z_i - \overline{Z}(\beta_0; t)d\widehat{M}_i(\beta_0; t). \tag{6}$$

Assume the covariates Z_i are uniformly bounded. For $i = 1, 2, \ldots, n$, we define

$$W_i = \int_{-\infty}^{\infty} \phi(\beta_0; t)\{Z_i - \overline{Z}(\beta_0; t)\}d\widehat{M}_i(\beta_0; t),$$

and summarize the following results as a lemma.

Lemma 2.1. *Under regularity conditions,* i) $n^{-1/2}\sum_{i=1}^{n} W_i \xrightarrow{\mathscr{D}} N(0, V_{\phi})$, *and* (ii) $n^{-1}\sum_{i=1}^{n} W_i W_i^T \to V_{\phi}$ *in probability.*

Thus, the empirical likelihood at true value β_0 is given

$$L(\beta_0) = \sup\left\{\prod_{i=1}^{n} p_i : \sum p_i = 1, \sum_{i=1}^{n} p_i W_i = 0, p_i \geq 0, i = 1, \ldots, n\right\}.$$

Let $p = (p_1, \ldots, p_n)^T$ be a vector of probabilities such that $\sum_{i=1}^{n} p_i = 1$, where $p_i \geq 0$, $i = 1, 2, \ldots, n$. Since $\prod_{i=1}^{n} p_i$ attains its maximum at $p_i = 1/n$, the empirical likelihood ratio at the true value β_0 is then

$$R(\beta_0) = \sup\left\{\prod_{i=1}^{n} np_i : \sum p_i = 1, \sum_{i=1}^{n} p_i W_i = 0, p_i \geq 0, i = 1, \ldots, n\right\}.$$

By using Lagrange multipliers, we know that $R(\beta_0)$ is maximized when

$$p_i = \frac{1}{n}\{1 + \lambda^T W_i\}^{-1}, \quad i = 1, \ldots, n,$$

where $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{W_i}{1 + \lambda^T W_i} = 0. \tag{7}$$

The value of λ may be found by numerical search (e.g., Newton–Raphson method), see the discussion in Owen (2001). Thus combining above equalities, we have

$$-2\log R(\beta_0) = 2\sum_{i=1}^{n} \log\{1 + \lambda^{\mathrm{T}} W_i\}, \tag{8}$$

where λ satisfies Eq. (7).

Hence, we establish the following theorem.

Theorem 2.1. *Assume the above conditions hold. Then $-2\log R(\beta_0)$ converges in distribution to χ_p^2, where χ_p^2 is a chi-squared distribution with degrees of freedom p.*

According to this theorem, an asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for β is

$$\mathcal{R}_3 = \{\beta : -2\log R(\beta) \leq \chi_p^2(\alpha)\}, \tag{9}$$

where $\chi_p^2(\alpha)$ is defined before.

3. Simulation Study

An extensive simulation is conducted to compare the relative performance of score test based procedure (SCORE), Wald-type procedure (WALD), model-based empirical likelihood procedure (MBEL), and model-free empirical likelihood procedure (MFEL). The SCORE is based on (4). The WALD is based on (3). The MBEL is based on Zhou (2005a). The MFEL is based on (9). We consider the extreme value distribution for error term. The covariate is uniformly distributed in $[-1, 1]$, and the censoring time is Uniform$[0, c]$, where c controls the censoring rate (CR). Corresponding to $\beta_0 = 2$, the censoring rates are approximately 15%, 30%, 45%, and 60%, respectively, which represent light censoring, moderate censoring, and heavy censoring rates, respectively. First, we choose $\phi = 1$ corresponding to the log-rank statistics. Second, we choose $\phi = S^{(0)}$ corresponding to the Gehan statistics. Simulations are carried out with a Fortran program that is available from the authors.

3.1. *Size Comparison*

We compare the four methods in terms of coverage probability. The sample size is set to be 10, 30, 50, 75, and 100, respectively. The simulation results for log-rank and Gehan estimating functions are tabulated in Tables 1 and 2, respectively. Each entry of the table is based on 10,000 simulated data sets. Note that "NA" in the table means the result is not applicable because of the failure of EM algorithm in the simulation.

From Tables 1 and 2, we find that at each nominal confidence level, the accuracy of coverage probabilities for all methods increases as the sample size n increases. As shown in the tables, all the methods work reasonably well with right coverage probabilities of 90%, 95%, and 99% when sample size is relatively large. The MFEL, MBEL, Wald-type, and score-based coverage probabilities tend to achieve the nominal levels with moderate sample sizes ($n = 50, 75, 100$). All the methods work well under censoring rates 15%, 30%, 45%, and 60%, respectively, for moderate sample.

Now, we compare the relative performance of two EL-related methods. Corresponding to log-rank statistics, from Table 1 we see that for small sample ($n = 10$, 30) the coverage probability of MBEL is more accurate than that of MFEL when censoring rate is 15% and 30%, and the coverage probability of MBEL is close to that of MFEL when censoring rate is 45% and 60%. That is, the advantage of MBEL disappears when censoring rate is high. Corresponding to Gehan statistics, from Table 2 we see that for small sample ($n = 10$, 30) the coverage probability of MFEL is more accurate than that of MBEL. Thus, the coverage probabilities for MFEL and MBEL are comparable in general.

For small sample size ($n = 10, 30$), the model-free empirical likelihood method apparently has relatively larger under-coverage, while the Wald-type method has better coverage for nominal level 90%, 95%. Note for nominal level 99%, when censoring rate increases the empirical likelihood becomes relatively better than Wald-type method. In particular, for higher censoring rate 60%, the empirical likelihood method has better coverage than the Wald-type method for nominal level 99%.

From Tables 1 and 2, we see that the coverage probability of the score test-based confidence region $\mathcal{R}_2$ based on (4) is the best among all the four methods. The score-based method outperforms EL-based methods remarkably. Its good performance may be expected from the fact the estimating function is rank based which is quite robust, stable, and is not sensitive to the outliers.

3.2. *Power Analysis*

Now, we do some power comparison for these tests. The null hypothesis is that β is 0.9, 1.4, 2.6, and 3.1, respectively, i.e., $H_0 : \beta_0 = 0.9$, or $H_0 : \beta_0 = 1.4$, or $H_0 : \beta_0 = 2.6$, or $H_0 : \beta_0 = 3.1$. The alternative hypothesis is $H_a : \beta_0 \neq 0.9$, or $H_a : \beta_0 \neq 1.4$, or $H_a : \beta_0 \neq 2.6$, or $H_a : \beta_0 \neq 3.1$. Let sample size n be 30, 50, and 100, respectively. The censoring rate is chosen to be approximately 15%, 30%, 45%, and 60%, respectively. In each case the powers are based on 10,000 samples and exact critical value 1.96^2 at nominal level $\alpha = 0.05$ is used. Data sets are simulated with $\beta_0 = 2.0$ and that the test H_0 is carried out, thus by counting the number of rejection of H_0. The corresponding power for these tests is given in Tables 3 and 4, respectively.

From the tables, the power decreases when censoring rate increases and the power increases when sample size increases. When the value of β is far away from $\beta_0 = 2$, the power increases and it is much easy to detect the H_a. Since from Tables 3 and 4, the power of test for $\beta = 0.9$ is the largest and the power of test for $\beta = 2.6$ is the smallest among the four values for fixed sample size n and censoring rate.

Table 1
Coverage probabilities for the regression parameter with log-rank statistics

		$1-\alpha=0.90$				$1-\alpha=0.95$				$1-\alpha=0.99$			
CR	n	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL
15%	10	.873	.882	.823	.765	.937	.944	.878	.827	.987	.986	.936	.903
	30	.893	.903	.880	.861	.948	.954	.935	.922	.989	.989	.985	.973
	50	.894	.897	.887	.877	.946	.948	.943	.931	.989	.989	.988	.981
	75	.895	.896	.895	.885	.948	.948	.947	.939	.989	.989	.989	.984
	100	.895	.897	.898	.888	.967	.948	.948	.942	.990	.989	.989	.987
30%	10	.884	.884	.814	.774	.940	.925	.870	.838	.989	.960	.937	.914
	30	.894	.894	.879	.871	.949	.945	.935	.926	.989	.987	.982	.979
	50	.899	.901	.888	.885	.951	.949	.943	.938	.989	.988	.988	.984
	75	.903	.905	.892	.894	.950	.951	.944	.946	.991	.989	.988	.988
	100	.902	.904	.901	.899	.951	.951	.951	.948	.991	.990	.990	.991
45%	10	.886	.849	NA	.754	.951	.892	NA	.817	.991	.943	NA	.906
	30	.901	.894	.869	.869	.952	.941	.924	.929	.992	.981	.975	.980
	50	.900	.896	.888	.886	.951	.943	.942	.940	.989	.985	.983	.984
	75	.902	.903	.898	.893	.952	.951	.949	.947	.991	.988	.989	.988
	100	.903	.906	.902	.898	.953	.949	.953	.949	.991	.989	.988	.991
60%	10	.896	.800	NA	.719	.951	.839	NA	.792	.988	.881	NA	.881
	30	.900	.891	.845	.857	.949	.933	.900	.915	.991	.972	.961	.972
	50	.896	.894	.876	.874	.948	.941	.932	.932	.989	.981	.978	.981
	75	.898	.895	.884	.887	.948	.943	.941	.939	.989	.983	.984	.985
	100	.904	.904	.895	.895	.953	.952	.947	.944	.991	.987	.986	.988

Table 2
Coverage probabilities for the regression parameter with Gehan statistics

		$1-\alpha=0.90$				$1-\alpha=0.95$				$1-\alpha=0.99$			
CR	n	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL
15%	10	.885	.873	.792	.787	.942	.932	.845	.854	.989	.979	.913	.926
	30	.897	.895	.876	.876	.948	.948	.928	.934	.990	.988	.978	.984
	50	.895	.894	.895	.884	.948	.947	.944	.941	.988	.988	.986	.986
	75	.901	.899	.894	.892	.949	.947	.946	.944	.988	.988	.988	.987
	100	.896	.894	.895	.892	.944	.943	.948	.943	.989	.988	.989	.987
	10	.892	.847	.734	.784	.946	.898	.792	.856	.993	.946	.869	.933
30%	30	.899	.885	.867	.873	.951	.936	.919	.929	.991	.982	.973	.982
	50	.901	.892	.885	.886	.952	.942	.938	.937	.990	.984	.983	.985
	75	.902	.898	.889	.891	.952	.947	.943	.944	.991	.987	.985	.989
	100	.900	.898	.898	.894	.950	.946	.952	.944	.991	.987	.988	.990
45%	10	.898	.812	NA	.760	.956	.854	NA	.842	.994	.907	NA	.931
	30	.907	.882	.845	.860	.956	.929	.902	.922	.993	.971	.956	.981
	50	.902	.885	.874	.878	.950	.935	.927	.931	.992	.978	.977	.982
	75	.899	.891	.887	.885	.951	.943	.940	.942	.992	.984	.984	.984
	100	.902	.897	.894	.892	.952	.945	.945	.944	.992	.985	.987	.989
60%	10	.905	.748	NA	.720	.954	.812	NA	.816	.991	.859	NA	.909
	30	.904	.860	.798	.825	.955	.908	.858	.894	.993	.949	.925	.970
	50	.898	.877	.849	.855	.951	.924	.908	.911	.991	.968	.962	.973
	75	.897	.885	.870	.871	.949	.933	.925	.927	.993	.976	.975	.981
	100	.902	.895	.881	.884	.953	.942	.936	.939	.992	.979	.982	.984

Table 3
Power analysis for the regression parameter with log-rank statistics

Null		$\beta = 0.9$				$\beta = 1.4$				$\beta = 2.6$				$\beta = 3.1$			
CR	n	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL
15%	30	.807	.798	.805	.848	.369	.378	.338	.435	.305	.203	.328	.336	.702	.572	.699	.722
	50	.953	.952	.953	.962	.559	.570	.554	.593	.470	.385	.467	.483	.916	.869	.881	.916
	100	.999	.999	1.00	.999	.840	.849	.848	.845	.791	.745	.775	.789	.998	.997	.994	.998
30%	30	.735	.771	.730	.772	.290	.342	.282	.340	.213	.113	.248	.244	.524	.333	.541	.558
	50	.918	.929	.924	.924	.456	.501	.451	.484	.324	.223	.354	.341	.773	.645	.739	.778
	100	.997	.997	.999	.997	.736	.759	.770	.745	.606	.519	.603	.607	.978	.963	.962	.976
45%	30	.612	.671	.638	.657	.219	.289	.231	.268	.159	.058	.199	.196	.374	.171	.416	.424
	50	.831	.864	.838	.842	.344	.406	.361	.370	.236	.125	.271	.259	.583	.396	.572	.597
	100	.984	.988	.990	.985	.591	.645	.628	.609	.434	.325	.437	.440	.898	.830	.858	.896
60%	30	.448	.565	.515	.510	.147	.243	.209	.205	.119	.037	.173	.165	.247	.069	.319	.318
	50	.663	.741	.693	.688	.233	.318	.251	.268	.167	.068	.195	.198	.393	.192	.419	.434
	100	.919	.945	.941	.926	.414	.476	.442	.439	.296	.179	.301	.303	.689	.537	.675	.692

Table 4
Power analysis for the regression parameter with Gehan statistics

Null		$\beta = 0.9$				$\beta = 1.4$				$\beta = 2.6$				$\beta = 3.1$			
CR	n	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL	SCORE	WALD	MBEL	MFEL
15%	30	.731	.736	.729	.764	.309	.330	.295	.353	.257	.211	.324	.296	.623	.539	.699	.654
	50	.909	.912	.913	.916	.482	.497	.467	.503	.399	.354	.467	.420	.857	.810	.881	.857
	100	.995	.995	.999	.995	.758	.766	.780	.766	.703	.668	.775	.708	.991	.988	.994	.990
30%	30	.652	.695	.671	.690	.253	.314	.250	.297	.185	.118	.249	.234	.458	.306	.541	.522
	50	.859	.875	.866	.869	.393	.449	.387	.422	.282	.207	.354	.313	.688	.562	.738	.701
	100	.988	.989	.993	.989	.657	.689	.677	.669	.523	.446	.603	.529	.947	.917	.962	.940
45%	30	.541	.626	.584	.595	.192	.291	.241	.253	.140	.065	.199	.205	.323	.152	.416	.428
	50	.768	.814	.781	.781	.296	.365	.311	.331	.201	.113	.251	.248	.504	.311	.567	.551
	100	.962	.969	.975	.965	.526	.578	.545	.542	.368	.267	.441	.384	.827	.714	.858	.819
60%	30	.397	.539	.495	.491	.134	.271	.251	.221	.104	.050	.178	.197	.215	.068	.318	.351
	50	.602	.699	.637	.641	.201	.307	.255	.261	.146	.065	.196	.214	.334	.134	.421	.436
	100	.880	.914	.903	.885	.368	.439	.377	.392	.249	.137	.301	.279	.602	.395	.670	.624

The power comparison of four tests are interesting. The powers for MBEL and MFEL tests are close in each case. For $H_0 : \beta_0 = 2.6$, or $H_0 : \beta_0 = 3.1$, the power for SCORE test is smaller than those for MBEL and MFEL tests when sample size n is 30, 50 and censoring rate is 45%, 60% and the power for the Wald-type test is the smallest among the four methods. In particular, when sample size is small or the censoring rate is heavy, the difference is very large. For $H_0 : \beta_0 = 0.9$, or $H_0 : \beta_0 = 1.4$, the power for the SCORE test is the smallest among the four methods when sample size is very small ($n = 30$) and the censoring rate is very heavy (CR = 60%).

3.3. *Conclusion*

Note that all these confidence intervals are test-based, i.e., constructed through the inversion of a test. We know that the shorter the average length of confidence interval, the better the confidence interval. Since Newton–Raphson algorithm does not work due to non-differentiable property of estimating equation with respect to β (cf. Huang, 2002), we could apply grid search to find confidence intervals for the four methods. We are not doing it in this article, largely due to the fact that the power analysis serves the same purpose. Thus, it is equivalent to do the power analysis for these four tests in Sec. 3.2. The larger the power of test, the better the test and more sensitive to detect the alternative hypothesis.

Before we summarize the comparison results of the methods, it is good to have in mind that for the best interval, the coverage probability needs to be as close as possible to the nominal confidence level while the average length needs to be the shortest. When coverage probabilities of two methods have close accuracies, we recommend the method which has shorter length, i.e., larger power. Based on coverage probability and power analysis, the model-based EL method and the model-free EL method are competitive methods. Overall, our simulation results suggest that these EL-based methods are not better than the standard approach based on the score test. The Wald-type method is the least favorable due to the small power.

4. Appendix

Proof of Theorem 1. Denote

$$M_i(\beta_0; t) = N_i(\beta_0; t) - \int_{-\infty}^{t} R_i(\beta_0; s)\lambda(s)ds.$$

We define

$$V_i = \int_{-\infty}^{\infty} \bar{\phi}(\beta_0; t)\{Z_i - \mu_Z(\beta_0; t)\}dM_i(\beta_0; t),$$

where $\mu_Z(\beta_0; t)$ is the limit of $\overline{Z}(\beta_0; t)$ as $n \to \infty$ and $\bar{\phi}(\beta_0; t)$ is the limit of $\phi(\beta_0; t)$. Then, it is clear that $E|V_i|^2 < \infty$. According to the proof of Lemma 3 in Owen (1990), we have $\max_{1\le i\le n}|V_i| = o_p(n^{1/2})$. By the martingale representation of V_i and W_i, we can prove that $\max_{1\le i\le n}|V_i - W_i| = o_p(n^{1/2})$. Then, we have

$$\max_{1\le i\le n}|W_i| = o_p(n^{1/2}), \quad \text{and} \tag{10}$$

$$\frac{1}{n}\sum_{i=1}^{n}|W_i|^3 = o_p(n^{1/2}). \tag{11}$$

Let $\lambda = \rho\theta$, where $\rho \geq 0$ and $|\theta| = 1$. Recall $\Gamma_n = 1/n \sum_{i=1}^n W_i W_i^T = V_\phi + o_p(1)$, where V_ϕ is the limit of $1/n \sum_{i=1}^n W_i W_i^T$. Let $\sigma_1 > 0$ be the smallest eigenvalue of V_ϕ. Then, $\theta \Gamma_n \theta \geq \sigma_1 + o_p(1)$. According to Lemma 2.1, $1/n \left| \sum_{i=1}^n W_i \right| = O_p(n^{-1/2})$.

By (10), the equations in (7) and the argument used in Owen (1990), we know that

$$|\lambda| = O_p(n^{-1/2}). \tag{12}$$

Consider a Taylor expansion to the right-hand side of (8),

$$-2\log R(\beta_0) = 2\sum_{i=1}^n \left\{ \lambda^T W_i - \frac{1}{2}(\lambda^T W_i)^2 \right\} + r_n, \tag{13}$$

where $|r_n| = O_p(1) \sum_{i=1}^n |\lambda^T W_i|^3$. Hence, by (11), $|r_n| = O_p(1)|\lambda|^3 \sum_{i=1}^n |W_i|^3 = o_p(1)$. Furthermore, since

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \frac{W_i}{1+\lambda^T W_i} &= \frac{1}{n}\sum_{i=1}^n W_i \left(1 - \lambda^T W_i + \frac{(\lambda^T W_i)^2}{1+\lambda^T W_i} \right) \\ &= \frac{1}{n}\sum_{i=1}^n W_i - \left(\frac{1}{n}\sum_{i=1}^n W_i W_i^T \right)\lambda + \frac{1}{n}\sum_{i=1}^n \frac{W_i(\lambda^T W_i)^2}{1+\lambda^T W_i} = 0, \end{aligned}$$

it follows that

$$\lambda = \left(\sum_{i=1}^n W_i W_i^T \right)^{-1} \sum_{i=1}^n W_i + o_p(1). \tag{14}$$

Similarly, we have

$$\sum_{i=1}^n \frac{\lambda^T W_i}{1+\lambda^T W_i} = \sum_{i=1}^n (\lambda^T W_i) - \sum_{i=1}^n (\lambda^T W_i)^2 + \sum_{i=1}^n \frac{(\lambda^T W_i)^3}{1+\lambda^T W_i} = 0. \tag{15}$$

Since

$$\sum_{i=1}^n \frac{(\lambda^T W_i)^3}{1+\lambda^T W_i} = o_p(1), \tag{16}$$

we know that $\sum_{i=1}^n (\lambda^T W_i)^2 = \sum_{i=1}^n \lambda^T W_i + o_p(1)$. Thus, the following is true:

$$\begin{aligned} -2\log R(\beta_0) &= \sum_{i=1}^n \lambda^T W_i + o_p(1) \\ &= \left(n^{-1/2} \sum_{i=1}^n W_i \right)^T \left(n^{-1} \sum_{i=1}^n W_i W_i^T \right)^{-1} \left(n^{-1/2} \sum_{i=1}^n W_i \right) + o_p(1) \\ &\xrightarrow{\mathscr{D}} \chi_p^2. \end{aligned}$$

Acknowledgments

The authors would like to thank Professor Mai Zhou for making available some computer programs. The authors also would like to thank the referees for their helpful comments which greatly improved the article. The research of Yichuan Zhao is partially supported by National Security Agency and Faculty Mentoring Grant, Georgia State University.

References

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

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

Fygenson, M., Ritov, Y. (1994). Montone estimating equations for censored data. *Annals of Statistics* 22:732–746.

Gehan, G. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* 52:203–223.

Huang, Y. (2002). Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association* 97:318–327.

Jin, Z., Lin, D. Y., Wei, L. J., Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* 90:341–353.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistical Probability Letters* 25:95–104.

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

Lin, D. Y., Wei, L. J., Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* 85:605–618.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its considerations. *Cancer Chemo Therapy Report* 50:163–170.

Owen, A. (1990). Empirical likelihood and confidence regions. *Annals of Statistics* 18:90–120.

Owen, A. (2001). *Empirical Likelihood*. New York: Chapman and Hall/CRC.

Parzen, M. I., Wei, L. J., Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* 81:341–350.

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

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

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

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals Statistics* 18:354–372.

Wei, L. J., Ying, Z., Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* 77:845–851.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* 21:76–99.

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

Zhou, M. (2005b). Empirical likelihood ratio with arbitrarily censored/truncated data by a modified EM algorithm. *Journal of Computer Graphical Statistics* 14:643–656.

Zhou, M., Li, G. (2004). Empirical likelihood analysis of the Buckley–James estimator. Department of Statistics, University of Kentucky, preprint.