Empirical Likelihood Analysis of the Buckley-James Estimator

Mai Zhou [1] and Gang Li [2]

University of Kentucky, Lexington, KY 40506
University of California, Los Angeles, CA 90095

Summary

The censored Accelerated Failure Time (AFT) model and the Buckley-James estimator (1979) is widely seen as an alternative to the popular Cox model when the assumption of proportional hazards is questionable. It performs well in many simulations and examples: Miller and Halpern (1982), Heller and Simonoff (1990, 1992), and Stare, Heinzl and Harrell (2000). The direct interpretation of the AFT model is more attractive than the Cox model, as D.R. Cox himself have pointed out. However, the application of the Buckley-James estimation was limited mainly due to its illusive variance estimation.

We use the empirical likelihood method (Owen (2001)) to derive a test (and thus confidence interval) based on the Buckley-James estimator of the regression coefficient. Standard chi square distribution is used to calculate the P-value and the construction of the confidence interval.

Simulations show that the chi square approximations of the log empirical likelihood ratio performs well.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.
Key Words and Phrases: Wilks theorem.

1. Introduction

The empirical likelihood method was first proposed by Thomas and Grunkmier (1975) to obtain better confidence intervals in connection with the Kaplan-Meier estimator. Owen (1988, 1990) and many others developed this into a general methodology. It has many desirable statistical properties, see the book of Owen (2001). Recently, the empirical likelihood ratio method has been shown to work with censored/truncated data. One of the nice features of the empirical likelihood method particularly appreciated in censored data analysis is that we can construct confidence intervals without estimating the variance of the statistic. Those variances can be very difficult to estimate, as in the situation of the Buckley-James estimator.

Cox proportional hazards regression model is very popular and successful in modeling covariate effects with censored data. However, there are many cases the proportional hazards model clearly

[1]Mai Zhou, Department of Statistics, University of Kentucky, Lexington, KY 40506 USA E-mail: mai@ms.uky.edu
[2]Gang Li, Department of Biostatistics, University of California, Los Angeles, CA 90095 USA E-mail: vli@ucla.edu

do not apply or are awkward to use. Other types of regression models that can handle censored data and are semi-parametric in nature, are needed. Cox himself said in an interview (Reid, 1994) "Of course, another issue is the physical or substantive basis for the proportional hazards model. I think that is one of its weaknesses, that accelerated life models are in many ways more appealing because of their quite direct physical interpretation, particularly in an engineering context". See also Wei (1992).

The Buckley-James estimator (1979) is an iterative estimator for the censored AFT model (regression model). The available of cheap, fast computer and ever-improving software in the last 10 to 15 years made the calculation of the Buckley-James estimator a routing business. But the variance estimator of the Buckley-James estimator remains very difficult. For example, the program `bj( )` within the `Design` library of Harrell (available for both S-plus and R) uses a variance estimation formula given by BJ's original paper which do not have rigorous justification and as Lai and Ying (1991) have pointed out this formula may not be correct. On the other hand, the variance given by Lai and Ying (1991) involves the density and the derivative of the density of the unknown distribution. The estimation of such functions can be highly unstable, unless we have a huge sample size.

Recent work of Li and Wang (2003) is a first attempt to use EL to tackle this problem. While their work is important and a step in the right direction, the application is hampered by the fact that the limiting distribution of their empirical likelihood ratio is not a standard chi square but a linear combination of several chi squares with coefficients depending on the unknown underlying distributions.

We propose in this paper a new type of EL testing procedure for the Buckley-James estimator where the likelihood is truly the censored likelihood and the limiting distribution is a regular chi square. Thus the P-value of the test and confidence interval can be obtained without estimating other quantities.

2. The Regression Model and the Empirical Likelihood

Consider the linear regression model

$$y_i = \beta^t x_i + \epsilon_i$$

where ϵ_i are iid with zero mean and finite variance.

The censored observations we have are

$$\tilde{y}_i = \min(y_i, c_i); \quad \delta_i = I_{[y_i \le c_i]}$$

where c_i are the censoring times, assumed independent of y_i given x_i. We assume x_i, a length q vector, is always observed.

For any candidate, b, of estimator of β, we define

$$e_i(b) = \tilde{y}_i - b^t x_i \ .$$

When $b = \beta$, we get the censored ϵ:

$$e_i(\beta) = \tilde{y}_i - \beta^t x_i := \tilde{\epsilon}_i$$

i.e. when $\delta_i = 1$ then $e_i(\beta) = \epsilon_i$ and when $\delta_i = 0$, $e_i(\beta) < \epsilon_i$.

Let us order the $e_i(b)$:

$$e_{(1)}(b) < \cdots < e_{(n)}(b)$$

and order the δ_i, x_i along with the e_i's. Notice this ordering is dependent on b. For simplicity, we assume the $e_i(b)$ is already ordered and save the notation: $e_i = e_{(i)}$.

Let $\hat{F}_{KM}(t, b)$ be the Kaplan-Meier estimator of F_ϵ based on the $e_i(b), \delta_i$. With the distribution $\hat{F}_{KM}(t, b)$, we can form a $(n \times n)$ weight matrix, M, as follows: if e_i is censored, i.e. $\delta_i = 0$ then let

$$m[i, j] = 0, \ j \le i \quad \text{and} \quad m[i, k] = \frac{\Delta \hat{F}_{KM}(e_k)}{1 - \hat{F}_{KM}(e_i)} \quad \text{for } k > i \ ;$$

if $\delta_i = 1$ then $m[i, i] = 1$, and $m[i, j] = 0; \ j \ne i$.

This matrix has the property: $\sum_j m[i, j] = 1$ for all i. Also M is an upper triangle matrix.

Let $\sum_i m[i, j] = n \times w_j$, then w_j, $j = 1, 2, \cdots, n$ is a probability with support on the uncensored e_i's. Since the Kaplan-Meier estimator is a self-consistent estimator, we have $w_j = \Delta \hat{F}_{KM}(e_j)$.

Actually the summation need only for those $i : i \leq j$. (triangle matrix).

The Buckley-James estimating equation is

$$0 = \sum_{i=1}^{n} \left(\delta_i x_i e_i(b) + (1-\delta_i) x_i \sum_{j:j>i} e_j(b) m[i,j] \right) \tag{1}$$

All n terms in the above summation are non-zero.

We can rewrite the Buckley-James estimating equation according to e_i

$$\begin{aligned} 0 &= \sum_i \delta_i e_i(b) \left[x_i + \sum_{k<i} m[k,i] x_k \right] && (2) \\ &= \sum_i \delta_i e_i(b) \left[x_i + \sum_{k<i, \delta_k=0} x_k \frac{\Delta F(e_i)}{1-F(e_k)} \right] && (3) \\ &= \sum_i \delta_i e_i(b) (m[,i] \cdot x) \ . && (4) \end{aligned}$$

The non zero term in the above summation now is the same as the number of uncensored e_i.

The equation to use with the censored data EL (to be defined in (6)) is:

$$0 = \sum_i e_i(b) \frac{m[,i] \cdot x}{n w_i} \ \delta_i p_i \tag{5}$$

where the $\cdot$ means inner product and $p_i = \Delta F(e_i)$, the jump of any distribution that has support on the uncensored e_i's.

The inner product, $m[,i] \cdot x$, above is

$$x_i + \sum_{k:k<i, \delta_k=0} m[k,i] x_k$$

since $m[i,i] = 1$ when $\delta_i = 1$.

The empirical likelihood (EL) for the $e_i(b)$ is defined as:

$$EL = \prod_{i=1}^{n} p_i^{\delta_i} (1 - \sum_{e_j \leq e_i} p_j)^{1-\delta_i} \ . \tag{6}$$

We are to find a distribution F or p_i's such that (a) it has support only on the un-censored e_i's; (b) it satisfy the estimating equation (5); and (c) among those F we shall find one that maximize the censored EL, (6).

Remark: When maximizing the censored EL with the equation (5), we only changing the p_i. The weight matrix M and w_i, are to remain unchanged for a fixed b.

Remark: Clearly, if $b = \hat{\beta}_{BJ}$ then $p_i = \Delta \hat{F}_{KM}(t, \hat{\beta}_{BJ})$ will satisfy the estimating equation (5), and maximizes the EL (6) among all CDF's. Where $\hat{\beta}_{BJ}$ is the Buckley-James estimator. Therefore, the confidence regions based on our Empirical Likelihood ratio will be "centered" at $\hat{\beta}_{BJ}$.

The computational problem of the constrained maximization, (5) and (6), with respect to p_i of the censored EL is the same as the one faced by the censored EL with mean constraint: $\sum f(t_i)p_i = \mu$. Here $f(t_i) = t_i(m[,i] \cdot x)/(nw_i)$ and $\mu = 0$. Zhou (2002) showed that this computation can be solved reliably by a modified EM algorithm, for sample sizes upward of 10,000 with ease. Basically the E-step is same as in Turnbull (1976) and the M-step is a weighted version of constrained maximization for uncensored data EL similar to the one used by Owen (1988). Zhou (2002) also proved that the modified EM algorithm is equivalent to the solution of the original constrained maximization of the censored EL. For details, please see Zhou (2002).

When $b = \beta$, the true parameter, then $y_i - \beta^t x_i$ are iid. The EL we use are the same as the censored EL based on iid right censored observations, used by Thomas and Grunkemeier (1976), Li (1995), Murphy and van der Varrt (1997) and Pan and Zhou (2001), among others.

The constraint equation we use, however, is slightly different to the mean constraint used by other people: the function $f(\cdot)$ we use depended on the data, and thus should be denoted by $f_n(\cdot)$:

$$\int f(t)dF(t) \quad v.s. \quad \int f_n(t)dF(t) \ .$$

We need the following generalization of the EL Wilks theorem for right censored data:

Theorem 1 *Suppose* (T_i, δ_i); $i = 1, 2, \cdots n$ *are iid right censored data where* $T_i = \min(X_i, C_i)$; $\delta_i = I_{[X_i \le C_i]}$. *Let* $P(X_i \le t) = F_0(t)$. *Define the censored empirical likelihood*

$$EL(F) = \prod_{i=1}^{n} p_i^{\delta_i}(1 - \sum_{T_j \le T_i} p_j)^{1-\delta_i} \ .$$

In addition, suppose for every n, $f_n(t)$ *is a random function but is predictable wrt* $\mathcal{F}_t$, *the usual counting process filtration (see for example Fleming and Harrington (1991)). Denote by* $M_n(t) = \sqrt{n}[\hat{F}_{KM}(t) - F_0(t)]/[1 - F_0(t)]$ *the* $\mathcal{F}_t$*-martingale associated with the Kaplan-Meier estimator* $\hat{F}_{KM}(t)$. *If the predictable function* $f_n(\cdot)$ *satisfy certain regularity conditions (so that* *the CLT for $\int f_n(t)dM_n(t)$ holds), and finally assume*

$$\int_{-\infty}^{\infty} f_n(t)dF_0(t) \equiv 0$$

then we have

$$-2\log\frac{\sup_F EL(F)}{EL(\hat{F}_{KM})} \xrightarrow{\mathcal{D}} \chi^2_{(1)}$$

where the numerator EL is maximized under the constraint

$$\sum \delta_i p_i f_n(T_i) = \int_{-\infty}^{\infty} f_n(t)dF(t) = 0 \ .$$

With the help of the following lemma, this theorem can be proved similarly to Pan and Zhou (2001), Pan (1997). We defer the proof of the theorem to appendix.

Lemma 1 *Let $g_n(t)$ be predictable functions such that $\int_{-\infty}^{\infty} g_n(t)dF_0(t) = 0$ and $g_n(t) \xrightarrow{\mathrm{P}} g(t)$ as $n \to \infty$ with $g(t)$ satisfy $\sigma^2_{KM}(g) < \infty$ (defined below), then we have*

$$\sqrt{n}\int_{-\infty}^{\infty} g_n(t)d\hat{F}_{KM}(t) = \sqrt{n}\sum g_n(T_i)\Delta\hat{F}_{KM}(T_i) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{KM}(g)) \ .$$

The asymptotic variance is given by

$$\sigma^2_{KM}(g) = \int_0^{\infty}\left\{g(x)[1-F_0(x)] - \int_{-\infty}^{x} g(s)dF_0(s)\right\}^2 \frac{dF_0(x)}{[1-F_0(x)]^2[1-G_0(x)]}. \tag{7}$$

Furthermore, the asymptotic variance can be consistently estimated by

$$\int_{-\infty}^{\infty}\left\{g_n(x)[1-\hat{F}_{KM}(x)] - \int_{-\infty}^{x} g_n(s)d\hat{F}_{KM}(s)\right\}^2 d\langle M_n(x)\rangle \ .$$

Proof: By the assumption $\int g_n(t)dF_0(t) = 0$ we have

$$\sqrt{n}\int_{-\infty}^{\infty} g_n(t)d\hat{F}_{KM}(t) = \sqrt{n}\int_{-\infty}^{\infty} g_n(t)d[\hat{F}_{KM}(t) - F_0(t)]$$

$$= \int g_n(t)d\{M_n(t)[1-F_0(t)]\} = \int M_n(t)g_n(t)d[1-F_0(t)] + \int g_n(t)[1-F_0(t)]dM_n(t) \ .$$

Integration by parts in the first integral above will give

$$= \int_{-\infty}^{\infty}\left\{g_n(t)[1-F_0(t)] - \int_t^{\infty} g_n(s)dF_0(s)\right\} dM_n(t) \ .$$

Use the fact that $\int g_n(t)dF_0(t) = 0$ again, we have

$$= \int_{-\infty}^{\infty} \left\{ g_n(t)[1 - F_0(t)] + \int_{-\infty}^{t} g_n(s)dF_0(s) \right\} dM_n(t) \ .$$

The integrand inside { } is clearly a predictable function in the above and thus the integration is also a martingale. By the CLT for martingales, it converges to a normal distribution with zero mean and a variance that can be consistently estimated by

$$\int_{-\infty}^{\infty} \left\{ g_n(t)[1 - F_0(t)] - \int_{-\infty}^{t} g_n(s)dF_0(s) \right\}^2 d \langle M_n(t) \rangle \ .$$

Replace $F_0(\cdot)$ by its consistent estimate $\hat{F}_{KM}(\cdot)$ gives the desired result. $\diamondsuit$

Lemma 2 *The weight function used in (5), $f_n(t_i) = t_i(m[, i] \cdot x)/(nw_i)$, is $\mathcal{F}_t$ predictable.*

PROOF: Notice if the Kaplan-Meier estimator jumps or not at t is not predictable but we are only concerned here with the *size* of the jump, if there is one. The size of the next jump of the Kaplan-Meier estimator can be computed from the history and thus is predictable. More specifically, the next jump size of the Kaplan-Meier estimator, at time t, if there is one, is equal to $1/n \times 1/(1 - \hat{G}(t-))$, where $\hat{G}$ is the Kaplan-Meier estimator of the censoring distribution.

Similarly we can infer from the history the portion of the jump, if there is one, that came from $t_j;\ t_j < t$. This proportion is precisely $m[j, i]/(nw_i)$. $\diamondsuit$

Armed with the generalized version of the ELT for right censored data, and also the Lemma 2, we have the Wilks theorem for the Buckley-James estimator:

Theorem 2 *When $\beta = \beta_0$, the residuals $e = \tilde{y} - \beta_0 x = \tilde{\epsilon}$ are iid censored observations and the estimating equation (5) can be written as*

$$E(eE^*(x|e = t)) \equiv 0$$

where E^* denotes the average over the $m[j, i]/(nw_i), j = 1, 2, \cdots n$. Thus, we have

$$-2 \log ELR(\beta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \ .$$

PROOF: Since x is independent of ϵ, $E(eE^*(x|e = t)) = E(x)E(e) \equiv 0$. $\diamondsuit$

3. Simulation for one dimensional β

We take the regression model

$$y_i = 2x_i + \epsilon_i$$

where x_i is uniform(0.5, 1.5); ϵ_i is uniform(-0.5, 0.5). We further take c_i to be $1 + 3.2\exp(1)$, where exp(1) represent a r.v. with standard exponential distribution. The test we carry out are based on the censored response observations: $\min(y_i, c_i)$, δ_i. Sample size is always 100.

The $-2\log$ empirical likelihood ratio are computed for each simulation run for the hypothesis $H_0 : \beta = 2$, which is true. The resulting Q-Q plot shows a good fit to the chi-square distribution with 1 degree of freedom.

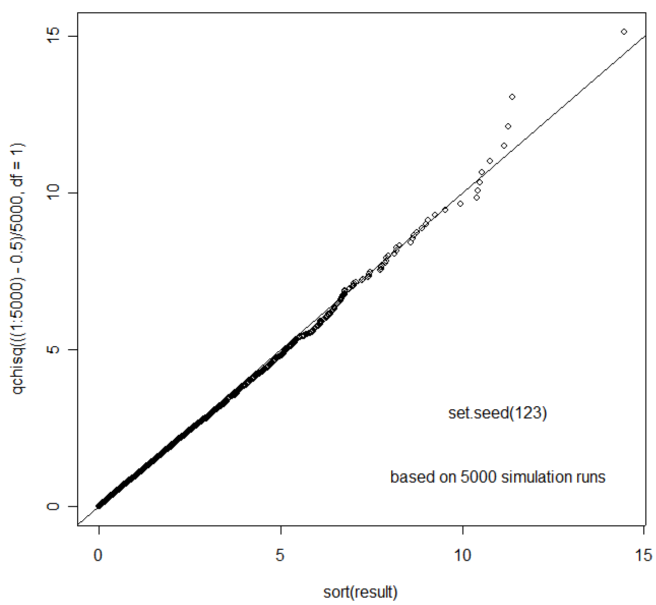

Figure 1: Q-Q plot of $-2\log ELR$, 5000 simulation run, sample size = 100.

In other simulations we used the regression model $y_i = x_i + \epsilon_i$ where ϵ_i are iid $N(0, sd = 0.5)$, x_i are iid $N(1, 0.5)$ and the censoring variables C_i are simulated from $N(\mu, sd = 4)$, with $\mu = -1.8, 1, 3.1$ and 6.1 respectively. This produces samples with censoring percentages equal to 75%, 50%, 30% and 10% approximately.

We used sample sizes of $n = 50, 100$ and 200. The coverage probabilities are based on 5000 simulation runs. The computation is done with the software R. These entry are in bold face. The relavant program `bjtest( )` can be obtained as part of the contributed package `emplik` from CRAN site:

`http://cran.us.r-project.org/`. The entry for ELEE and ELSD are from Li and Wang (2003).

		Nominal level = 90%			Nominal level = 95%		
Sample	Censoring	Coverage probability			Coverage probability		
size	rate	ELEE	ELSD	BJ	ELEE	ELSD	BJ
50	0.75	0.94	0.79	**0.8420**	0.98	0.85	**0.9042**
100	0.75	0.94	0.83	**0.8818**	0.97	0.90	**0.9344**
200	0.75	0.92	0.87	**0.8928**	0.96	0.92	**0.9438**
50	0.50	0.94	0.84	**0.8838**	0.98	0.91	**0.9324**
100	0.50	0.94	0.88	**0.8926**	0.97	0.95	**0.9414**
200	0.50	0.92	0.91	**0.8952**	0.97	0.95	**0.9482**
50	0.30	0.94	0.87	**0.8866**	0.97	0.92	**0.9374**
100	0.30	0.93	0.89	**0.8936**	0.97	0.93	**0.9472**
200	0.30	0.92	0.91	**0.8922**	0.97	0.95	**0.9468**
50	0.10	0.95	0.85	**0.8924**	0.98	0.93	**0.9406**
100	0.10	0.93	0.91	**0.8888**	0.97	0.94	**0.9404**
200	0.10	0.93	0.90	**0.8810**	0.97	0.94	**0.9458**

Table 1. Comparison of empirical likelihood confidence intervals for β: ELEE is the Li and Wang (2003) empirical likelihood method based on estimating equations and ELSD refers to the empirical likelihood method of Jing and Qin (2002) and Li and Wang (2003) based on synthetic data. BJ is the method proposed in this paper.

Note: Each entry is based on 2000 Monte Carol samples. Except the bold entry is based on 5000 simulations.

We remind readers that besides the performance differences of the empirical likelihood confidence intervals, the estimator $\hat{\beta}$ itself of Buckley-James and those based on synthetic data are also different.

4. Some Extensions

In this section we briefly discuss some extension of the Buckley-James estimator and the EL analysis with them.

4.1 M-estimator

For complete data the regression M-estimator is defined as the minimizer of $\sum \rho(y_i - \beta^t x_i)$ or the solution to the equation

$$\sum x_i \psi(y_i - \beta^t x_i) = 0$$

where $\psi(t) = \frac{d\rho(t)}{dt}$. Usually we assume ψ is monotone. Let $e_i = y_i - \beta^t x_i$.

With right censored data, the Buckley-James estimating equation is

$$0 = \sum_{i=1}^{n} \left(\delta_i x_i \psi(y_i - \beta^t x_i) + (1-\delta_i) x_i \sum_{j:\ e_j > e_i} \psi(y_j - \beta^t x_j) m[i,j] \right) .$$

A rewriting of the estimating equation according to e_i gives

$$0 = \sum_i \delta_i \psi(y_i - \beta^t x_i) \left[x_i + \sum_{k:\ e_k < e_i} m[k,i] x_k \right] .$$

In the EL analysis of the censored Buckley-James regression M-estimator, the definition of the censored empirical likelihood remains unchanged as in (6). The constraint (or estimating equation) to be used with the empirical likelihood is

$$0 = \sum_{i=1}^{n} \psi(y_i - \beta^t x_i) \frac{x_i + \sum_{k:\ e_k < e_i} m[k,i] x_k}{n w_i} \delta_i p_i .$$

Similar results as in the least squares estimator can be obtained for the regression M-estimator. We omit the details here.

5. Examples

In this section we illustrate the EL analysis of the Buckley-James estimator with the Stanford Heart Transplant data. Following Miller and Halpern (1982), we use only 152 cases.

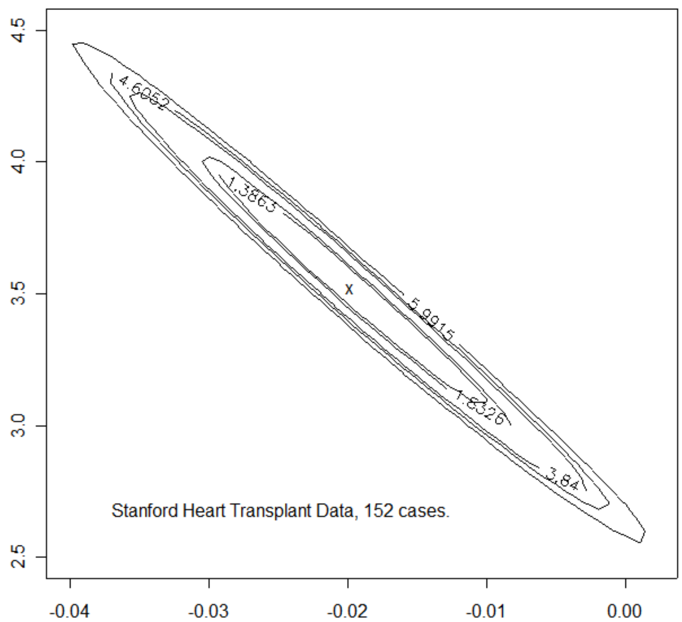

Figure 2: Contour plot for the -2logELR, Stanford Heart Transplant Data, 152 cases.

The Buckley-James estimator of the $(\hat{\beta}_0, \hat{\beta}_1)$ was marked by an `X` on the plot. From the plot we see that the contours are fairly symmetrical and elliptically shaped, indicating that the normal approximation is pretty good for the Buckley-James estimator here.

From the plot we see that the estimator $\hat{\beta}_0$ is strongly negatively correlated with $\hat{\beta}_1$. The 95% confidence interval for the β_1 alone is approximately $[-0.0357, -0.0028]$, the 95% confidence interval for β_0 alone is approximately $[2.755, 4.255]$. These are obtained as the left(right, upper or lower) most point of the contour with level 3.84. They are approximate values because we used a coarse grid points to produce the contour plot, and thus interpolation was used in the plot.

From the `bj( )` function from the `Design` library of F. Harrell, the following results are obtained:

```
> bj(Surv(log10(time), status)~age, data=stanford5, link="identity")
Buckley-James Censored Data Regression

bj(formula = Surv(log10(time), status) ~ age, data = stanford5, link = "identity")

       Obs     Events       d.f. error d.f.      sigma
```

```
      152         97           1          95     0.6796

              Value Std. Error      Z  Pr(>|Z|)
Intercept   3.52696   0.299123 11.79 4.344e-32
age        -0.01990   0.006632 -3.00 2.700e-03
```

Our confidence intervals are slightly wider than the ones obtained by the Wald confidence interval using the standard error estimator given by the function `bj( )`. We remind readers that the standard error estimator produced by `bj( )` has no theoretical justifications.

References

Buckley, J. J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66** 429-36.

Fleming and Harrington (1991). *Counting Processes and Survival Analysis.* Wiley.

Gentleman, R. and Ihaka, R. (1996). R: A Language for data analysis and graphics. *J. of Computational and Graphical Statistics*, **5**, 299-314.

Heller, G. and Simonoff, J.S. (1990). A comparison of estimators for regression with a censored response variable. *Biometrics*, **77**, 515-520.

Heller, G. and Simonoff, J.S. (1992). Prediction in censored survival data: a comparison of the proportional hazards and linear regression models. *Biometrics*, **48**, 101-115.

Lai T.L. and Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist*, **19**, 1370-402.

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

Li, G. and Wang, Qi-Hua (2003). Empirical likelihood regression analysis for right censored data. *Statistica Sinica*, **13**, 51-68.

Miller, RG. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69** 521-31.

Hillis SL. (1995). Residual plots for the censored data linear regression model. *Stat. in Med.*; 14: 2023-2036.

Murphy, S. and van der Vaart, A. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471-1509.

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

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

Owen, A. (2001). *Empirical likelihood.* Chapman & Hall, London.

Pan, X. R. (1997). Empirical Likelihood Ratio Method for Censored Data. Ph. D. dissertation, Department of Statistics, University of Kentucky.

Pan, X.R. and Zhou, M. (1999). Using one parameter sub-family of distributions in empirical likelihood with censored data. *J. Statist. Planning and Infer.* **75**, 379-392.

Reid, N. (1994). A conversation with Sir David Cox. *Statist. Sci.* **9**, 439-455.

Stare, J., Heinzl, H., and Harrell, F. (2000). On the Use of Buckley and James Least Squares Regression for Survival Data. *New Approaches in Applied Statistics*, Ferligoj, A. and Mrvar, A. (Editors).

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

Turnbull, B. (1976), The empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSS* **B**, 290-295.

Wei, L.J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statist. Med.* **11**, 1871-1879.

Zhou, M. (2002). Empirical likelihood ratio with arbitrarily censored/truncated data by a modified EM algorithm. Tech Report. University of Kentucky.

Appendix

We briefly outline the proof of Theorem 1 here. First of all we define a class of functions

$$\mathcal{H}_g^{F_0} = \left\{ h \mid \; h \text{ is left continuous}, \int h^2 dF_0 < \infty, g_n(t)h(t) \geq 0 \; a.s. \right\}.$$

Furthermore we define a one-parameter family of distribution functions

$$\mathcal{A}_h^{F_0} = \left\{ F_\lambda(t) \mid F_\lambda(t) = \sum_{i:T_i \leq t} \Delta F_\lambda(T_i) \right\},$$

where

$$\Delta F_\lambda(T_i) = \Delta \hat{F}_{KM}(T_i) \times \frac{1}{1+\lambda h(T_i)} \times \frac{1}{C(\lambda)}, \quad i = 1, 2, ..., n, \tag{8}$$

and $C(\lambda)$ is just a normalizing constant

$$C(\lambda) = \sum_{i=1}^n \frac{\Delta \hat{F}_{KM}(T_i)}{1+\lambda h(T_i)}.$$

The parameter λ is well defined in a neighborhood of zero and for $\lambda = 0$, we get back the the Kaplan-Meier: $F_{\lambda=0} = \hat{F}_{KM}$. Within this family of distributions, there is only one that satisfy the constraint equation

$$\int g_n(t) dF_\lambda(t) = 1/C(\lambda) \sum_{i=1}^n \Delta \hat{F}_{KM}(T_i) \frac{g_n(T_i)}{1+\lambda h(T_i)} = 0. \tag{9}$$

We denote the parameter for this unique distribution as λ_0.

Finally we define a class of profile empirical likelihood ratio functions as follows:

$$\mathcal{R}_h^{F_0}(\theta) = \left\{ \frac{L(F_{\lambda_0})}{L(\hat{F}_{KM})} \mid F \in \mathcal{A}_h^{F_0} \right\}.$$

Lemma A *Assume all the conditions in Lemma 1. Then, as $n \to \infty$, (1) $\lambda_0 = O_p(n^{-1/2})$, (2) $n\lambda_0^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times \dfrac{\sigma^2_{KM}(g)}{(\int ghdF_0)^2}$.*

PROOF: (outline for proof of (2)). Expanding (9), we have

$$\begin{aligned} 0 &= \sum_{i=1}^n \Delta \hat{F}_{KM}(T_i) \frac{g_n(T_i)}{1+\lambda_0 h(T_i)} \\ &= \sum_{i=1}^n g_n(T_i) \Delta \hat{F}_{KM}(T_i) - \lambda_0 \sum_{i=1}^n g_n(T_i) h(T_i) \Delta \hat{F}_{KM}(T_i) \\ &\qquad + \lambda_0^2 \sum_{i=1}^n \frac{g_n(T_i) h^2(T_i)}{1+\lambda_0 h(T_i)} \Delta \hat{F}_{KM}(T_i), \end{aligned}$$

from there we have

$$\lambda_0 = \frac{\sum_{i=1}^n g_n(T_i) \Delta \hat{F}_{KM}(T_i)}{\sum_{i=1}^n g_n(T_i) h(T_i) \Delta \hat{F}_{KM}(T_i)} + o_p(n^{-1/2}).$$

By Lemma 1, as $n \to \infty$,

$$\sqrt{n}(\sum_{i=1}^{n} g_n(T_i)\Delta \hat{F}_{KM}(T_i)) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{KM}(g)),$$

and

$$\sum_{i=1}^{n} g_n(T_i)h(T_i)\Delta \hat{F}_{KM}(T_i) \xrightarrow{\mathrm{P}} \int_0^\infty g(x)h(x)dF_0(x).$$

By Slutsky's theorem $n\lambda_0^2 \to \chi^2_{(1)} \times a_h$ in distribution as $n \to \infty$, where

$$a_h = \sigma^2_{KM}(g)/(\int_0^\infty g(x)h(x)dF_0(x))^2. \tag{10}$$

$\diamondsuit$

Theorem A *If the conditions in Lemma A hold, then, as* $n \to \infty$

$$-2\log \mathcal{R}^{F_0}_{h_0}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times r_h,$$

where

$$r_h = \frac{\sigma^2_{KM}(g) \times c_h}{(\int ghdF_0)^2},$$

and

$$c_h = \int_0^\infty h^2(x)(1-G_0(x))dF_0(x) + \int_0^\infty \frac{(\int_x^\infty h(s)dF_0(s))^2}{1-F_0(x)}dG_0(x) - \left(\int_0^\infty h(x)dF_0(x)\right)^2.$$

Furthermore, $\inf_h r_h = 1$.

PROOF: Define

$$f(\lambda) = \log \prod_{i=1}^{n} (\Delta F_\lambda(T_i))^{\delta_i}(1-F_\lambda(T_i))^{1-\delta_i}, \tag{11}$$

where $|\lambda| \le |\lambda_0|$ and $F \in \mathcal{A}^{F_0}_h$. From the definition we can see that

$$C(0) = 1 \quad \text{and} \quad f(0) = \log \prod_{i=1}^{n} (\Delta \hat{F}_{KM}(T_i))^{\delta}(1-\hat{F}_{KM}(T_i))^{1-\delta_i} = L(\hat{F}_{KM}).$$

By Lemma A, $\lambda_0 = O_p(n^{-1/2})$ where λ_0 is the root of (9). Hence we can apply Taylor's expansion for $f(\lambda_0)$:

$$f(\lambda_0) = f(0) + \lambda_0 f^{'}(0) + \frac{\lambda_0^2}{2} f^{''}(0) + \frac{\lambda_0^3}{3!} f^{'''}(\xi), \qquad |\xi| \le |\lambda_0|.$$

Substituting (8) in (11),

$$\begin{aligned} f(\lambda) \quad = \quad & \sum_{i=1}^{n} \delta_i \log \Delta \hat{F}_{KM}(T_i) - \sum_{i=1}^{n} \delta_i \log(1+\lambda h(T_i)) - n \log \left(\sum_{i=1}^{n} \frac{\Delta \hat{F}_{KM}(T_i)}{1+\lambda h(T_i)} \right) \\ & + \sum_{i=1}^{n} (1-\delta_i) \log \left(\sum_{j:T_j > T_i} \frac{\Delta \hat{F}_{KM}(T_j)}{1+\lambda h(T_j)} \right). \end{aligned}$$

Some tedious but straight forward calculation of the derivatives show that the first derivative

$$f^{'}(0) = -\sum_{i=1}^{n}\delta_i h(T_i) + n\sum_{i=1}^{n} h(T_i)\Delta\hat{F}_{KM}(T_i) - \sum_{i=1}^{n}(1-\delta_i)\frac{\sum_{j:T_j>T_i} h(T_i)\Delta\hat{F}_{KM}(T_j)}{1-\hat{F}_{KM}(T_j)} = 0\ ;$$

and the second derivative of f with respect to λ, evaluated at $\lambda = 0$ is

$$\begin{aligned}
&f^{''}(0)\\
&= n(\sum_{i=1}^{n} h(T_i)\Delta\hat{F}_{KM}(T_i))^2 - n\sum_{i=1}^{n} h^2(T_i)\Delta\hat{F}_{KM}(T_i)\\
&\quad + \sum_{i=1}^{n}(1-\delta_i)\frac{\sum_{j:T_j>T_i} h^2(T_i)\Delta\hat{F}_{KM}(T_i)}{1-\hat{F}_{KM}(T_i)} - \sum_{i=1}^{n}(1-\delta_i)\frac{(\sum_{j:T_j>T_i} h(T_i)\Delta\hat{F}_{KM}(T_i))^2}{(1-\hat{F}_{KM}(T_i))^2},
\end{aligned}$$

where, by Theorem 2.2 of Zhou (1986, p.6), as $n \to \infty$,

$$\sum_{i=1}^{n} h(T_i)\Delta\hat{F}_{KM}(T_i) \xrightarrow{\mathrm{P}} \int_0^\infty h(x)dF_0(x),$$

$$\sum_{i=1}^{n} h^2(T_i)\Delta\hat{F}_{KM}(T_i) \xrightarrow{\mathrm{P}} \int_0^\infty h^2(x)dF_0(x),$$

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{(\sum_{j:T_j>T_i} h(T_i)\Delta\hat{F}_{KM}(T_i))^2}{(1-\hat{F}_{KM}(T_i))^2} \xrightarrow{\mathrm{P}} \int_0^\infty \frac{\left(\int_x^\infty h(s)dF_0(s)\right)^2}{1-F_0(x)}dG_0(x).$$

Hence

$$-\frac{1}{n}f^{''}(0) \xrightarrow{\mathrm{P}} c_h, \tag{12}$$

where

$$c_h = \int_0^\infty h^2(x)(1-G_0(x))dF_0(x) + \int_0^\infty \frac{\left(\int_x^\infty h(s)dF_0(s)\right)^2}{1-F_0(x)}dG_0(x) - \left(\int_0^\infty h(x)dF_0(x)\right)^2. \tag{13}$$

Finally by similar calculations we can show that the third derivative of f evaluated at ξ is

$$f^{'''}(\xi) = o_p(n^{2/3}). \tag{14}$$

Now observe

$$\begin{aligned}
-2\log\mathcal{R}_h^{F_0}(\theta_0) &= 2\left(f(0) - f(0) - \lambda_0 f^{'}(0) - \frac{\lambda_0^2}{2}f^{''}(0) - \frac{\lambda_0^3}{3!}f^{'''}(\xi)\right)\\
&= -\lambda_0^2 f^{''}(0) - \frac{\lambda_0^3}{3}f^{'''}(\xi).
\end{aligned}$$

By Lemma A , (12), (14), and Slutsky theorem, we obtain

$$-2\log\mathcal{R}_h^{F_0}(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \times r_h,$$

where $r_h = \dfrac{\sigma^2_{KM}(g) \times c_h}{(\int ghdF_0)^2}$, and c_h is defined by (13).

We now proof the infimum of the constant r_h over h is one. First we notice that

$$\frac{\left(\int h^2(1-G)dF + \int \frac{[\int_t^\infty h(s)dF(s)]^2}{1-F(t)} dG(t) - [\int hdF]^2\right)}{\left(\int ghdF\right)^2}$$

is precisely the information defined by van der Vaart (1991), as i_α in his (4.1).

The infimum of i_α over all one-dimensional submodels is called "efficient Fisher information". And in this case (right censored observations), the reciprocal of it is given by the last equation on p. 193 of van der Vaart (1991), (as the lower bound for the asymptotic variance of estimating $\int gdF$):

$$\inf i_\alpha = \frac{1}{||\beta||_F^2} = \left(\int \frac{(R_{\tilde{\chi}_F})^2}{1-G} dF\right)^{-1} .$$

Lastly, we notice that $\int g_n d\hat{F}_{KM}$ is an efficient estimate and therefore we can easily check

$$\text{Asy Var}\left(\int gd\hat{F}_{KM}\right) = \int \frac{(R_{\tilde{\chi}_F})^2}{1-G} dF .$$

Therefore, $\inf_h r_h = 1$. $\diamondsuit$