October 20, 2007

FULL LIKELIHOOD INFERENCES IN THE COX MODEL

By Jian-Jian Ren[1] and Mai Zhou[2]
University of Central Florida and University of Kentucky

Abstract

We use the empirical likelihood approach to parameterize the full likelihood function of the Cox model via baseline distribution F_0 instead of the usual baseline hazard parameterization. After explicitly profiling out nuisance parameter F_0, the profile likelihood function for regression parameter β_0 is obtained, and the maximum likelihood estimator (MLE) for (β_0, F_0) is derived. The relation between the MLE and Cox's partial likelihood estimator for β_0 is made clear by showing that Taylor's expansion gives Cox's partial likelihood estimating function as the leading term of the profile estimating function induced by the full likelihood function. We show that the log full-likelihood ratio has an asymptotic chi-squared distribution, while the simulation studies indicate that for small or moderate sample sizes, the MLE performs favorably over Cox's partial likelihood estimator. Moreover, we present a real dataset example, where our full-likelihood ratio test and Cox's partial likelihood ratio test lead to statistically different conclusions.

1 Introduction

Since Cox (1972), the following Cox's regression model has become one of the most widely used tools in analyzing survival data:

$$\lambda(t; z) = \lambda_0(t) \exp(z^{\mathrm{T}}\beta), \tag{1}$$

where Z is a p-dimensional vector of covariates, β is the regression parameter, and $\lambda(t;\, z)$ is the conditional hazard function of random variable (r.v.) X given $Z = z$ with $\lambda_0(t)$ as an arbitrary baseline hazard function. Suppose that $(X_1, Z_1), \cdots, (X_n, Z_n)$ is a random sample of (X, Z), and the actually observed censored survival data are

$$(V_1, \delta_1, Z_1), (V_2, \delta_2, Z_2), \cdots, (V_n, \delta_n, Z_n), \tag{2}$$

where $V_i = \min\{X_i, Y_i\}, \delta_i = I\{X_i \le Y_i\}$, and Y_i is the right censoring variable that has a distribution function (d.f.) F_Y and is independent of (X_i, Z_i) or independent of

[1]Research supported in part by NSF grant DMS-0604488.
[2]Research supported in part by NSF grant DMS-0604920.
AMS 2000 subject classifications. 62N02.
Key words and phrases. Right censored data, empirical likelihood, maximum likelihood estimator, partial likelihood.

X_i given $Z = Z_i$. Then, letting β_0 be the true value of β in (1), Cox's *partial likelihood estimator* $\hat{\beta}_c$ for β_0 is given by the solution of equations (Tsiatis, 1981):

$$(3) \qquad \varphi_n(\beta) \equiv n^{-1} \sum_{i=1}^{n} \delta_i \Big(Z_i - \frac{\sum_{j=1}^{n} I\{V_j \geq V_i\} Z_j \exp(Z_j^{\mathrm{T}} \beta)}{\sum_{j=1}^{n} I\{V_j \geq V_i\} \exp(Z_j^{\mathrm{T}} \beta)} \Big) = 0.$$

In the past few decades, $\hat{\beta}_c$ has been considered as the standard estimate for β_0 in statistical literature. Efficiency properties of $\hat{\beta}_c$ were discussed by Efron (1977) and Oakes (1977). In particular, Efron (1977) examined the complete likelihood function that is parameterized through baseline hazard function $\lambda_0(t)$ in (1), and showed that Cox's partial likelihood function contains *nearly all* of the information about β_0, and $\hat{\beta}_c$ is asymptotically efficient. Using the counting process approach, the books by Fleming and Harrington (1991), and Andersen, Borgan, Gill and Keiding (1993) give a complete treatment of asymptotic theory and include many relevant references. We also refer to Cox and Oakes (1984), Therneau and Grambsch (2000), Kalbfleisch and Prentice (2002) for discussions and references on developments of the Cox model.

However, as pointed out in Cox and Oakes (1984; page 123), the efficiency results on $\hat{\beta}_c$ are only asymptotic, and for finite samples the loss in precision from using the partial likelihood can be rather substantial. It is well known and confirmed clearly by our simulation results (some of which are presented in Section 3) that the loss of efficiency can occur when, among other possible situations, the sample size is small or moderate, or β_0 is far from 0. It is also well known that in medical clinical trials, the sample size of survival data is often small or moderate. With these in mind, a natural question would be: Does the actual maximum likelihood estimator (MLE) for β_0 (i.e., the MLE based on the complete or full likelihood) perform better for small or moderate samples? We do not know the answer to this question because up to now the actual MLE has not been given in the literature.

Using Poisson process arguments and parameterizing via baseline hazard $\lambda_0(t)$, Efron (1977) showed that the complete or full likelihood function can be expressed as the product of Cox's partial likelihood function and a factor which involves both β and observed data; see equation (3.10) of Efron (1977). This means that for finite samples, the inference based on the partial likelihood is not based on *all* the observed data in the sense that the partial likelihood is *not* the likelihood of observed sample (2); see discussions on page 559 of Efron (1977). But, in Efron's formula it is not obvious how to profile out nuisance parameter $\lambda_0(t)$ in order to obtain the actual MLE for β_0.

In this article, we use the empirical likelihood approach (Owen, 1988) to parameterize the full likelihood function of the Cox model (1) through F_0, which is the baseline d.f.

corresponding to baseline hazard function $\lambda_0(t)$. After explicitly profiling out nuisance parameter F_0, the (profile) likelihood function for β_0 is obtained and is not too much more complicated than the partial likelihood function; thus the actual MLE for β_0 can easily be computed. Note that the key to achieving our results here is the combination of utilizing the *Lehmann family* properties and our current understanding of the empirical likelihood techniques. Although Cox's partial likelihood has been carefully studied in the past 35 years, the *Lehmann family* properties, which are equivalent to the Cox model assumption (1), have not been used in the literature to parameterize the full likelihood function for the Cox model.

Based on our full likelihood function for (β_0, F_0) under the Cox model (1) with observed sample (2), Section 2 derives the MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0), where the relation between $\hat{\beta}_n$ and $\hat{\beta}_c$ is made clear by showing that Taylor's expansion gives Cox's partial likelihood estimating function $\varphi_n(\beta)$ as the leading term of the profile estimating function $\psi_n(\beta)$ induced by the full likelihood function. Section 2 also shows that the log full-likelihood ratio has an asymptotic chi-squared distribution. Section 3 discusses computational issues and treatment of ties, and presents some simulation results which show that the MLE $\hat{\beta}_n$ performs favorably over $\hat{\beta}_c$ for small or moderate sample sizes, especially when β_0 is away from 0. In Section 4, we discuss a real dataset example, where our full-likelihood ratio test and Cox's partial likelihood ratio test lead to statistically different conclusions.

The findings in this article suggest that the MLE is preferred over Cox's partial likelihood estimator when sample size n is small or moderate. While this should not be a surprise since the MLE is based on *all* the observed data in the sense that it is based on the likelihood of observed sample (2), further studies can help better understand the estimation bias for finite samples which will be considered in a separate paper. Our other appealing findings here include: (a) Wilk's theorem holds for the log full-likelihood ratio of β_0; (b) the full likelihood function leads to the MLE jointly for (β_0, F_0); (c) the computation for the MLE $\hat{\beta}_n$ is only slightly more complicated than Cox's partial likelihood estimator $\hat{\beta}_c$; (d) our method can be extended to deal with other types of censored data. This last point is of particular interest because it is well known that the counting process approach is applicable to right censored data, but not complicated types of censored data, such as doubly censored data (Chang and Yang, 1987; Gu and Zhang, 1993), interval censored data (Groeneboom and Wellner, 1992), etc.

While the main focus of this paper is the estimation of β_0 with small or moderate sample size n, it is worth noting that the MLE $\hat{F}_n$ for baseline distribution F_0 does not require any extension or approximation of the continuous proportional hazard model to discrete data; rather it is based on a full likelihood function with possible candidate d.f. that assigns all its probability mass to observations V_i's and interval $(V_{(n)}, \infty)$. In contrast, there have been several competing methods for estimation of $F_0(t)$ or the cumulative baseline hazard function $\Lambda_0(t)$ that require the use of discrete logistic model (Cox, 1972), or grouping continuous model (Kalbfleisch and Prentice, 1973), or discretizing continuous proportional hazard model to have approximated MLE (Breslow, 1974) in the context of counting process (Andersen and Gill, 1982), etc.. For detailed discussions and more references, we refer to Andersen, Borgan, Gill and Keiding (1993; Section IV.1.5) and Kalbfleisch and Prentice (2002; page 143).

2 Maximum Likelihood Estimators

For simplicity of presentation, in this section we consider the case when covariate Z is a scaler rather than a vector, i.e., $p = 1$ in (1). Since the generalization of our results to multivariate case is straightforward, the results for case with $p > 1$ are summarized at the end of this section.

To parameterize the full likelihood function via F_0, we notice that under the assumption of the Cox model (1), each X_i has a d.f. that satisfies

$$\bar{F}(t \,|\, Z_i) = [\bar{F}_0(t)]^{c_i} \quad \Leftrightarrow \quad f(t \,|\, Z_i) = c_i f_0(t)[\bar{F}_0(t)]^{c_i - 1}, \tag{4}$$

where $c_i = \exp(Z_i\beta), \bar{F}_0(t) = [1 - F_0(t)]$ and $F(t \,|\, Z_i)$ is the conditional d.f. of X_i given $Z = Z_i$, while $f(t \,|\, Z_i)$ and $f_0(t)$ are the density functions of $F(t \,|\, Z_i)$ and $F_0(t)$, respectively. As the usual empirical likelihood treatment for continuous d.f.'s, we let

$$\begin{aligned} &P\{X = t \,|\, Z = z\} = dF(t \,|\, z) = F(t \,|\, z) - F(t- \,|\, z), \\ &P\{Y = t\} = dF_Y(t) = F_Y(t) - F_Y(t-), \\ &dF_0(t) = F_0(t) - F_0(t-), \end{aligned}$$

and treat $f(t \,|\, z)\, dt = dF(t \,|\, z)$ and $f_0(t)\, dt = dF_0(t)$. Then, under the Cox model (1) with data (2), the likelihood function of (V_i, δ_i) given $Z = Z_i$ is given by

$$\prod_{i=1}^{n} P\{V = V_i, \delta = \delta_i \,|\, Z = Z_i\} = \prod_{i=1}^{n} \Big(\bar{F}_Y(V_i)\, dF(V_i \,|\, Z_i)\Big)^{\delta_i} \Big(dF_Y(V_i)\bar{F}(V_i \,|\, Z_i)\Big)^{1-\delta_i},$$

which under (4) is proportional to

$$\prod_{i=1}^{n} [F(V_i \,|\, Z_i) - F(V_i- \,|\, Z_i)]^{\delta_i} [\bar{F}(V_i \,|\, Z_i)]^{1-\delta_i} = \prod_{i=1}^{n} \Big(c_i[F_0(V_i) - F_0(V_i-)]\Big)^{\delta_i} \Big(\bar{F}_0(V_i)\Big)^{c_i - \delta_i}.$$

Hence, if, without loss of generality, we assume that there are no ties among V_i's and assume that $V_1 < \cdots < V_n$ with $p_i = F(V_i) - F(V_i-)$, the full likelihood function for (β_0, F_0) in Cox model (1) with right censored data (2) is given by

$$L(\beta, F) = \prod_{i=1}^{n} (c_i p_i)^{\delta_i} (\textstyle\sum_{j=i+1}^{n+1} p_j)^{c_i - \delta_i}, \tag{5}$$

where $F(x) = \sum_{i=1}^n p_i I\{V_i \le x\}$ satisfying $\sum_{i=1}^{n+1} p_i = 1$ with $0 \le p_{n+1} \le 1$.

Denoting $d_i = c_i + \cdots + c_n$, we show in the Appendix that for any fixed value β satisfying $c_n \ge 1$, likelihood function $L(\beta, F)$ is maximized by:

$$1 - \hat{F}_n(t) = \prod_{V_i \le t} \frac{d_i - \delta_i}{d_i}. \tag{6}$$

In (5), we replace F by $\hat{F}_n$, then from the proof of (6) given in the Appendix (see (A.1)), we obtain the following profile likelihood function for β_0:

$$l(\beta) = \prod_{i=1}^{n} \Big(\frac{c_i}{d_i}\Big)^{\delta_i} \Big(\frac{d_i - \delta_i}{d_i}\Big)^{d_i - \delta_i}. \tag{7}$$

Thus, the MLE for β_0 is given by the solution $\hat{\beta}_n$ which maximizes the value of $l(\beta)$, and consequently $\hat{F}_n$ in (6) with β replaced by $\hat{\beta}_n$ is the MLE for F_0.

Differentiating $\log l(\beta)$, algebra shows that $\hat{\beta}_n$ should be a solution of equation

$$\psi_n(\beta) \equiv n^{-1} \sum_{i=1}^{n} \delta_i \Big(Z_i + e_i \log \frac{d_i - 1}{d_i} \Big) = 0, \tag{8}$$

where $e_i = Z_i c_i + \cdots + Z_n c_n$, and due to (7), $\log 0$ is set to 0 whenever it occurs. The Newton-Raphson method can be used to compute $\hat{\beta}_n$.

Remark 1. *On Condition* $c_n \ge 1$: Throughout this section so far, all arguments require $c_n \ge 1$ for any fixed β, which ensures a well-defined (6) (i.e., all terms on the right-hand side of equation are between 0 and 1) because $d_i > c_n \ge 1$ for all $1 \le i < n$. Note that the requirement of $c_n \ge 1$ for any positive or negative β is equivalent to requiring $Z_n = 0$. Thus, in practice and for the rest of this paper, the natural way to handle this is to adjust Z_i to $\tilde{Z}_i = Z_i - Z_n, 1 \le i \le n$, which rewrites model (1) as $\lambda(t; Z_i) = \lambda_{\beta,n}(t) \exp(\tilde{Z}_i \beta)$ with $\lambda_{\beta,n}(t) = \lambda_0(t) \exp(Z_n \beta)$. In (5)-(8), we replace Z_i and c_i by $\tilde{Z}_i$ and $\tilde{c}_i = \exp(\tilde{Z}_i \beta)$, respectively, then we have that $\tilde{c}_n \equiv 1$ for any β; the solution of (8) gives the MLE for β_0, still denoted as $\hat{\beta}_n$; and the resulting estimator in

(6), still denoted by $\bar{\hat{F}}_n$, is the MLE for $[\bar{F}_0(t)]^{e^{Z_n\beta}}$, thus the MLE for $\bar{F}_0(t)$ is given by $[\bar{\hat{F}}_n(t)]^{e^{-Z_n\hat{\beta}_n}}$. Our extensive simulation studies show that such a treatment on condition $c_n \geq 1$ gives excellent performance on the resulting MLE $\hat{\beta}_n$ and the Newton-Raphson algorithm. Finally, it should be noted that adjusting Z_i to $\tilde{Z}_i = Z_i - Z_n, 1 \leq i \leq n$, does not change the Cox's partial likelihood function, thus does not affect $\hat{\beta}_c$.

Interestingly, by Taylor's expansion we show in the Appendix that the MLE $\hat{\beta}_n$ is linked with Cox's partial likelihood estimator $\hat{\beta}_c$ by the following:

$$\psi_n(\beta) = \varphi_n(\beta) + O_p\Big(\frac{\log n}{n}\Big), \tag{9}$$

where $\varphi_n(\beta)$ is the partial likelihood estimating function in (3). Further, Wilk's theorem on the log full-likelihood ratio is established below with proof given in the Appendix.

Theorem 1. *Assume (9) and assume the regularity conditions on Cox model (1) (Andersen and Gill, 1982). Then,* $R_0 = -2\log[l(\beta_0)/l(\hat{\beta}_n)]$ *converges in distribution to a chi-squared distribution with 1 degree of freedom as* $n \to \infty$.

Remark 2. *On* p *-Dimensional Covariate* Z_i: If Z_i and β are p-dimensional vectors with $p > 1$ in (1)-(2), with minor modifications on the derivations and the proofs we have that (4)-(9) hold with $c_i = \exp(Z_i^\top \beta)$ and e_i as p-dimensional vectors, which imply that (8) has p equations. Moreover, a minor modified proof of Theorem 1 shows that R_0 converges in distribution to a chi-squared distribution with p degree of freedom.

3 Simulations

This section first presents some simulation results to compare the MLE $\hat{\beta}_n$ with Cox's partial likelihood estimator $\hat{\beta}_c$ for the case without ties among V_i's in (2). Then, we discuss how to handle ties among V_i's, and present some simulation results to compare $\hat{\beta}_n$ with Efron's estimator $\hat{\beta}_E$. In all our simulation studies, $\hat{\beta}_n$ is calculated using the Newton-Raphson method with $\hat{\beta}_c$ or $\hat{\beta}_E$ as the initial value for the algorithm. Routines in FORTRAN for computing $\hat{\beta}_n$ are available from the authors.

Without Ties Among V_i's in (2):

Let Exp(μ) represent the exponential distribution with mean μ, and U$(0,1)$ the uniform distribution on $(0,1)$. In our simulation studies, we consider $F_Y = \text{Exp}(2)$ as the d.f. of the right censoring variable Y_i, $F_Z = \text{U}(0,1)$ as the d.f. of Z, and $F_{X|Z} = \text{Exp}(e^{-Z\beta_0})$ as the conditional d.f. of X given Z; thus (X, Z) satisfies the Cox model (1) with regression parameter β_0 and baseline d.f. $F_0 = \text{Exp}(1)$. For each case of

$\beta_0 = 1, 0, -1$, we generate 1000 samples with sample size $n = 15, 20, 30, 50$, respectively, and for each n Table 1 includes the simulation average of $\hat{\beta}_c$ and $\hat{\beta}_n$ with the simulation standard deviation (s.d.) given in the parenthesis next to them, respectively. The censoring percentage in each case is also reported in Table 1.

Table 1. Comparison between $\hat{\beta}_c$ and $\hat{\beta}_n$

Parameter	$\beta_0 = 1$		$\beta_0 = 0$		$\beta_0 = -1$	
Sample Size	Ave. $\hat{\beta}_c$	Ave. $\hat{\beta}_n$	Ave. $\hat{\beta}_c$	Ave. $\hat{\beta}_n$	Ave. $\hat{\beta}_c$	Ave. $\hat{\beta}_n$
$n = 15$	1.145 (1.42)	1.042 (1.36)	0.020 (1.47)	0.016 (1.41)	-1.120 (1.72)	-1.062 (1.66)
$n = 20$	1.116 (1.13)	1.035 (1.09)	0.052 (1.23)	0.047 (1.16)	-1.112 (1.44)	-1.060 (1.39)
$n = 30$	1.081 (0.85)	1.018 (0.83)	0.013 (0.93)	0.009 (0.89)	-1.066 (1.04)	-1.018 (1.02)
$n = 50$	1.036 (0.64)	0.988 (0.63)	0.003 (0.68)	-0.001 (0.66)	-1.020 (0.74)	-0.986 (0.74)
Censoring %	23.7%		33.4%		45.4%	

Table 1 clearly shows that the MLE $\hat{\beta}_n$ performs better than Cox's partial likelihood estimator $\hat{\beta}_c$ for small or moderate sample sizes when β_0 is away from 0. For instance, the loss in precision for $\beta_0 = \pm 1$ with, say, $n = 15$ is reflected by the simulation Mean Square Error (MSE). Simple calculation gives that when $\beta_0 = 1$, the simulation MSE is 2.037 and 1.851 for $\hat{\beta}_c$ and $\hat{\beta}_n$, respectively, yielding $1.851/2.037 = 90.9\%$ (such ratio is 92.2% for $n = 20$), while when $\beta_0 = -1$, the simulation MSE is 2.973 and 2.759 for $\hat{\beta}_c$ and $\hat{\beta}_n$, respectively, yielding $2.759/2.973 = 92.8\%$ (such ratio is 92.7% for $n = 20$). The loss in precision for $\hat{\beta}_c$ shows even more obviously when we use $\beta_0 = \pm 2, \pm 3, \cdots$ in simulation studies of Table 1. To illustrate, we include results for $\beta_0 = -2$ with sample size $n = 15$ in Table 2, where Relative Bias is $|(\hat{\beta} - \beta_0)/\beta_0|$, Relative MSE is $E[(\hat{\beta} - \beta_0)/\beta_0]^2$ and the censoring variable is still Exp(2). Note that the ratio of simulation relative MSE for $\hat{\beta}_n$ and $\hat{\beta}_c$ is $8.969/33.658 = 26.6\%$ in Table 2.

Table 2. Comparison between $\hat{\beta}_c$ and $\hat{\beta}_n$

	$\beta_0 = -2$, $n = 15$, [Censoring Percentage] = 57.0%		
Estimator	Simulation Mean (s.d.)	Simulation Relative Bias	Simulation Relative MSE
$\hat{\beta}_c$	-3.403 (11.518)	0.702	33.658
$\hat{\beta}_n$	-2.828 (5.932)	0.414	8.969

Finally, although not presented here, our simulation studies also show that according to Remark 1, the MLE $\hat{F}_n$ given in (6) provides a very good estimate for $1 - [\bar{F}_0(t)]^{e^{Z_n \beta}}$.

With Ties Among V_i's in (2):

Let $W_1 < \cdots < W_m$ be all the distinct observations of $V_1 \leq \cdots \leq V_n$, where $m < n$, and for those tied V_j's, the uncensored V_j's are ranked ahead of the censored V_j's. If we have, say, $V_1 = V_2 = V_3 = W_1$ with $\delta_1 = \delta_2 = 1, \delta_3 = 0$, then by Efron's estimation (see pages 48-49; Therneau and Grambsch, 2000) d_i's in Cox's partial likelihood function are modified as $d_1 = c_1 + c_2 + \cdots + c_n$, $d_2 = (c_1 + c_2)/2 + c_3 + \cdots + c_n$, $d_3 = c_3 + \cdots + c_n$, etc., which give Efron's estimator $\hat{\beta}_E$. Applying these modified d_i's in (7)-(8), the MLE for β_0 when V_i's have ties is given by the solution of (8), still denoted by $\hat{\beta}_n$.

Some simulation results are presented in Tables 3-4 to compare the MLE $\hat{\beta}_n$ with $\hat{\beta}_E$. In these simulation studies, we consider $n = 15$, $F_Y =$Exp(2), and $F_{X|Z} = \text{Exp}(e^{-Z\beta_0})$, and we create ties among V_i's as follows: compute $t_k = V_1 + \frac{k}{n}(V_n - V_1)$ for $0 \leq k \leq n+1$, and set $V_i = t_{k+1}$ if $V_i \in [t_k, t_{k+1})$, which represents rounding errors in practice that cause tied values among V_i's. Table 3 includes the simulation results based on 1000 samples with $F_Z = \text{U}(0,1)$ for $\beta_0 = -2$ and $\beta_0 = 2$, respectively, and reports the average number m of distinct V_i's. Table 4 includes results of the same simulation studies with $F_Z =$Exp(1) for $\beta_0 = -0.75$ and $\beta_0 = 0.75$, respectively.

Table 3. Comparison between $\hat{\boldsymbol{\beta}}_{\boldsymbol{E}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$

$F_Z = \text{U}(0,1)$	Estimator	Simul. Mean (s.d.)	Rel. Bias	Rel. MSE	Ave. m
$\beta_0 = -2$	$\hat{\beta}_E$	-3.025 (9.044)	0.513	20.711	8.3
Censoring: 57.0%	$\hat{\beta}_n$	-2.662 (5.541)	0.331	7.785	8.3
$\beta_0 = 2$	$\hat{\beta}_E$	2.167 (1.349)	0.084	0.462	7.8
Censoring: 16.8%	$\hat{\beta}_n$	2.013 (1.329)	0.007	0.442	7.8

Table 4. Comparison between $\hat{\boldsymbol{\beta}}_{\boldsymbol{E}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$

$F_Z =$Exp(1)	Estimator	Simul. Mean (s.d.)	Rel. Bias	Rel. MSE	Ave. m
$\beta_0 = -0.75$	$\hat{\beta}_E$	-1.115 (2.490)	0.487	11.259	8.2
Censoring: 50.4%	$\hat{\beta}_n$	-1.029 (1.602)	0.372	4.701	8.2
$\beta_0 = 0.75$	$\hat{\beta}_E$	0.749 (0.435)	0.001	0.336	7.9
Censoring: 20.9%	$\hat{\beta}_n$	0.713 (0.428)	0.049	0.328	7.9

Overall, Tables 3-4 show that the MLE $\hat{\beta}_n$ performs favorably. In particular, note that the ratio of simulation relative MSE for $\hat{\beta}_n$ and $\hat{\beta}_E$ is $7.785/20.711 = 37.59\%$ for $\beta_0 = -2$ in Table 3, while such ratio is $4.701/11.259 = 41.8\%$ for $\beta_0 = -0.75$ in Table 4.

4 Data Example

We consider the *Stanford Heart Transplant* data set (Escobar and Meeker, 1992; it is available in R library with file name 'stanford2'), where Z is the age of a patient and X is the survival time subject to right censoring. To see the smaller sample performance, we use observations number $76 - 100$ and observations number $50 - 100$, respectively, to test $\mathrm{H}_0 : \beta = 0$ vs. $\mathrm{H}_1 : \beta \neq 0$ using Wald test, the *partial likelihood ratio* (PLR) test and our *full likelihood ratio* (FLR) test according to Theorem 1. The results are summarized in Table 5, which show that PLR test and FLR test can lead to statistically different conclusions for smaller sample size n.

Table 5. Stanford Heart Transplant Data

	n	Censored Obs.	$\hat{\beta}_c$	$\hat{\beta}_n$	Wald Test	PLR Test	FLR Test
Obs. used: $76 - 100$	25	8	0.367	0.397	0.063	0.056	0.038
Obs. used: $50 - 100$	51	23	0.153	0.149	0.050	0.045	0.049

Appendix

***Proof of* (6):** Let $a_i = p_i/b_i$ and $b_i = \sum_{j=i}^{n+1} p_j$. Then, we have that $b_1 = 1$, $b_{n+1} = p_{n+1}$, $b_{i+1} = (b_i - p_i)$, $(1 - a_i) = b_{i+1}/b_i$, and algebra can rewrite (5) as

$$L(\beta, F) = \prod_{i=1}^{n} (c_i p_i)^{\delta_i} (b_i - p_i)^{c_i - \delta_i} = \prod_{i=1}^{n} (c_i a_i)^{\delta_i} (1 - a_i)^{d_i - \delta_i}. \tag{A.1}$$

From the 1st and 2nd partial derivatives of $\log L$ with respect to a_i's, we know that the solution of equations $\partial(\log L)/\partial a_i = 0$ is given by $\hat{a}_i = 1/d_i, 1 \leq i \leq n$, and it maximizes $L(\beta, F)$ under condition $c_n \geq 1$. Hence, (6) follows from noting that $\bar{\hat{F}}_n(t) = \prod_{V_i \leq t}(1 - \hat{a}_i)$ and that condition $c_n \geq 1$ implies all $0 \leq \hat{a}_i \leq 1$. $\square$

***Proof of* (9):** We give the proof assuming that $c_n = 1$ (based on Remark 1), $|\beta| \leq M_\beta < \infty$ and Z has a finite support. From Taylor's expansion, we have in (8),

$$\begin{aligned} \psi_n(\beta) &= n^{-1} \sum_{i=1}^{n} \delta_i Z_i - n^{-1} \sum_{i=1}^{n-1} \delta_i e_i \Big(\frac{1}{d_i} + \frac{1}{2\xi_i^2} \Big) + \frac{e_n}{n} \log \frac{c_n - \delta_n}{c_n} \\ &= \varphi_n(\beta) + n^{-1} \delta_n Z_n - \tfrac{1}{2} R_n = \varphi_n(\beta) + O_p(n^{-1}) - \tfrac{1}{2} R_n, \end{aligned} \tag{A.2}$$

where $R_n = n^{-1}\sum_{i=1}^{n-1}(\delta_i e_i)/\xi_i^2$ with ξ_i being between d_i and $(d_i - \delta_i)$, and we have

$$\begin{aligned} |R_n| &\le \frac{1}{n}\sum_{i=1}^{n-1}\frac{|e_i|}{(d_i-1)^2} = \frac{1}{n}\sum_{i=1}^{n-1}\frac{|e_i|}{d_i^2(1-1/d_i)^2} \le \frac{1}{n}\sum_{i=1}^{n-1}\frac{|e_i|}{d_i^2}\Big(1+\frac{1}{c_{n-1}}\Big)^2 \\ &\le O_p(n^{-1})\max_{1\le i\le n}|Z_i|\sum_{i=1}^{n-1}\frac{\exp(-m_i(\beta))}{\exp(Z_i\beta - m_i(\beta)) + \cdots + \exp(Z_n\beta - m_i(\beta))} \\ &\le O_p(n^{-1})\sum_{i=1}^{n-1}\frac{1}{n-i+1} = O_p\Big(\frac{\log n}{n}\Big) \end{aligned} \tag{A.3}$$

for $m_i(\beta) = \min\{Z_j\beta \,|\, i \le j \le n\}$. The proof follows from (A.2)-(A.3). $\square$

Proof of Theorem 1: Applying Taylor's expansion on $\log l(\beta_0)$ at point $\hat\beta_n$, we have that from $\psi_n(\beta) = n^{-1}\frac{d}{d\beta}\log l(\beta)$ in (8) and $\psi_n(\hat\beta_n) = 0$,

$$R_0 = -n\psi_n'(\xi)(\beta_0 - \hat\beta_n)^2 = -\psi_n'(\xi)[\sqrt{n}(\hat\beta_n - \beta_0)]^2, \tag{A.4}$$

where ξ is between $\hat\beta_n$ and β_0. From (9), we know that $\sqrt{n}(\hat\beta_n - \beta_0)$ and $\sqrt{n}(\hat\beta_c - \beta_0)$ have the same limiting distribution $\mathrm{N}(0,\sigma^2)$ for some constant $0 < \sigma^2 < \infty$. It suffices to show that $-\psi_n'(\xi)$ converges to $1/\sigma^2$ in probability as $n \to \infty$.

From Taylor's expansion and $c_n = 1$, we have that in (9)

$$\begin{aligned} -\psi_n'(\beta) &= -\, n^{-1}\sum_{i=1}^{n-1}\Big(e_i'\log\frac{d_i-\delta_i}{d_i} + \frac{\delta_i e_i^2}{d_i(d_i-\delta_i)}\Big) \\ &= -\, n^{-1}\sum_{i=1}^{n-1}\Big\{e_i'\Big(-\frac{\delta_i}{d_i} - \frac{\delta_i}{2\xi_i^2}\Big) + \frac{\delta_i e_i^2}{d_i}\Big(\frac{1}{d_i} + \frac{\delta_i}{\eta_i^2}\Big)\Big\} \\ &= -\, \varphi_n'(\beta) + \tfrac{1}{2}R_{1,n} - R_{2,n}, \end{aligned} \tag{A.5}$$

where $R_{1,n} = n^{-1}\sum_{i=1}^{n-1}\delta_i e_i'/\xi_i^2$ and $R_{2,n} = n^{-1}\sum_{i=1}^{n-1}\delta_i e_i^2/(d_i\eta_i^2)$ with ξ_i and η_i being between d_i and $(d_i - \delta_i)$. Applying the argument in (A.3) to $R_{1,n}$ and $R_{2,n}$, respectively, we obtain $-\psi_n'(\beta) = -\varphi_n'(\beta) + O_p((\log n)/n)$. The proof follows from the fact that $-\varphi_n'(\beta)$ is the negative second derivative of the log of Cox's partial likelihood, and $-\varphi_n(\beta_0)$ converges to $1/\sigma^2$ in probability as $n \to \infty$; see Andersen and Gill (1982). $\square$

REFERENCES

Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N.(1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Chang, M.N. and Yang, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15**, 1536-1547.

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

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* Chapman & Hall, London.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557-65.

Escobar, L.A. and Meeker, W.Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507-528.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Groeneboom, P. and Wellner, J.A. (1992). *Information Bounds and Nonparametric Maximum likelihood Estimation.* Birkhäuser Verlag.

Gu, M.G. and Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611-624.

Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267-278.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons.

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441-448.

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

Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data, Extending the Cox Model.* Springer.

Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108.

Department of Mathematics
University of Central Florida
Orlando, Florida 32816
E-mail: jren@mail.ucf.edu

Department of Statistics
University of Kentucky
Lexington, Kentucky 40506
E-mail: mai@ms.uky.edu