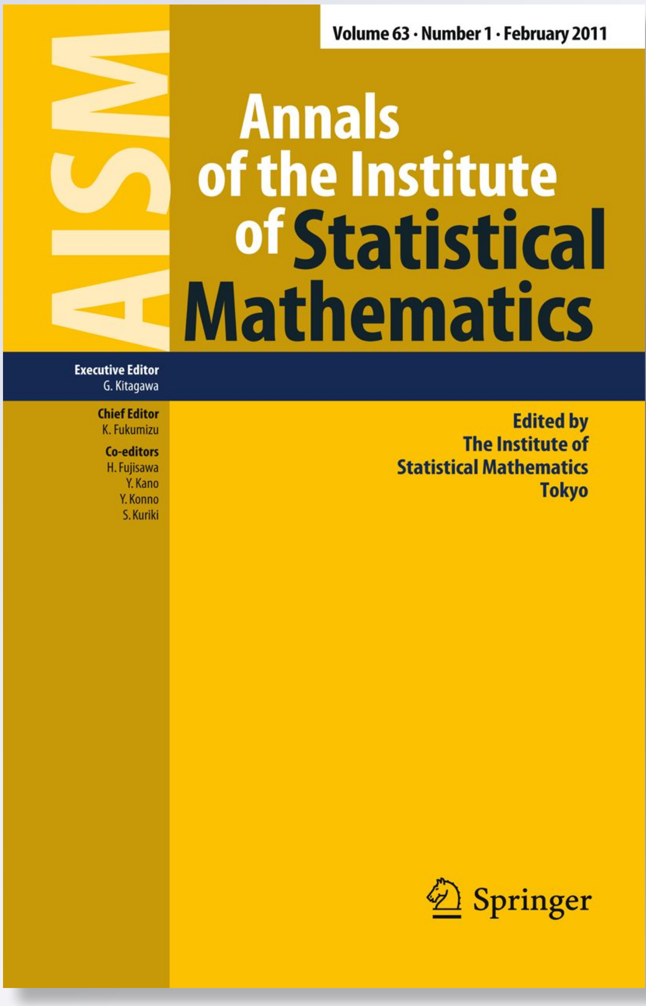
Full likelihood inferences in the Cox model: an empirical likelihood approach

Jian-Jian Ren & Mai Zhou

Annals of the Institute of Statistical Mathematics

ISSN 0020-3157
Volume 63
Number 5

Ann Inst Stat Math (2011)
63:1005-1018
DOI 10.1007/s10463-010-0272-y



Ann Inst Stat Math (2011) 63:1005–1018
DOI 10.1007/s10463-010-0272-y

Full likelihood inferences in the Cox model: an empirical likelihood approach

Jian-Jian Ren · Mai Zhou

Received: 22 September 2008 / Revised: 30 March 2009 / Published online: 18 February 2010

Abstract For the regression parameter β_0 in the Cox model, there have been several estimators constructed based on various types of approximated likelihood, but none of them has demonstrated small-sample advantage over Cox's partial likelihood estimator. In this article, we derive the full likelihood function for (β_0, F_0), where F_0 is the baseline distribution in the Cox model. Using the empirical likelihood parameterization, we explicitly profile out nuisance parameter F_0 to obtain the full-profile likelihood function for β_0 and the maximum likelihood estimator (MLE) for (β_0, F_0). 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 full-profile likelihood estimating 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. In a real dataset example, our full likelihood ratio test and Cox's partial likelihood ratio test lead to statistically different conclusions.

Keywords Right censored data · Empirical likelihood · Maximum likelihood estimator · Partial likelihood · Profile likelihood

J.-J. Ren's research was partially supported by NSF Grants DMS-0604488 and DMS-0905772 and M. Zhou's research was partially supported by NSF Grant DMS-0604920.

J.-J. Ren (✉)
Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA
e-mail: jren@mail.ucf.edu

M. Zhou
Department of Statistics, University of Kentucky, Lexington, KY 40506, USA
e-mail: mai@ms.uky.edu

1 Introduction

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

$$\lambda(t; z) = \lambda_0(t) \exp(z^{\mathrm{T}} \beta_0), \tag{1}$$

where Z is a p-dimensional vector of covariates, β_0 is the regression parameter, and $\lambda(t; z)$ is the conditional hazard function of continuous random variable (r.v.) X given $Z = z$ with $\lambda_0(t)$ as an arbitrary baseline hazard function. Suppose that $(X_1, Z_1), \ldots, (X_n, Z_n)$ is a random sample of (X, Z), and the actually observed censored survival data are

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

where $V_i = \min\{X_i, Y_i\}$, $\delta_i = I\{X_i \leq Y_i\}$, and Y_i is the right censoring variable with distribution function (d.f.) F_Y and is independent of (X_i, Z_i) or independent of X_i given $Z = Z_i$. Then, Cox's *partial likelihood estimator* $\hat{\beta}_c$ for β_0 is given by the solution of equations:

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

see Tsiatis (1981). 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 et al. (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 more discussions and references on the 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 Sect. 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. For instance, the numerical results presented in Table 1 of Kay (1979) show up to 18% loss of efficiency for $\beta_0 = 0.5$ with rather large sample size $n = 1000$. It is also well known that in medical clinical trials, the sample size of survival data is often small or moderate. In literature, there have been some estimators for β_0 constructed based on various types of approximated likelihood, viz., the likelihood by grouping continuous model Kalbfleisch and Prentice (1973), the likelihood by discretizing the continuous Cox model Bailey (1984), etc. But none of these estimators has demonstrated small-sample advantage over Cox's partial likelihood estimator $\hat{\beta}_c$. 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 exact 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 parameterization 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 derive the full likelihood function for (β_0, F_0) in Cox model (1) with survival data (2), where F_0 is the baseline distribution corresponding to baseline hazard function $\lambda_0(t)$. Using the empirical likelihood parameterization (Owen 1988), we explicitly profile out nuisance parameter F_0 to obtain the full-profile likelihood function for β_0 as well as the actual MLE for β_0.

Note that the key to achieving our results here is the combination of utilizing the *Lehmann family* properties and the empirical likelihood techniques. Although Cox's partial likelihood has been carefully studied in the past 35 years, the *Lehmann family* properties have not been used to derived the exact full likelihood function for the continuous Cox model (1). The empirical likelihoods have had a long history in survival analysis. Since Owen (1988), empirical likelihood has been developed as a powerful nonparametric inference approach, and many authors have applied it in survival analysis; see Owen (2001) and a review paper by Li et al. (2005) for detailed descriptions and many references on this topic. As reviewed in Li et al. (2005) for works on the Cox model with right censored data, Pan (1997) obtained an approximated empirical likelihood function for (β_0, Λ_0) in the Cox model (1), where Λ_0 is the cumulative hazard function of λ_0. By discretizing Λ_0, she showed that nuisance parameter Λ_0 is profiled out via the Breslow estimator (Breslow 1974), and the resulting profile likelihood function for β_0 is the Cox's partial likelihood. This result was also obtained in Murphy and van der Vaart (2000), who referred their approach as empirical likelihood due to the use of point masses on the hazard function. However, one of the unsatisfactory features of the Breslow estimator is that the estimated hazard masses can exceed 1 (Kalbfleisch and Prentice 2002, pages 116–117), in turn, the corresponding estimator for $\bar{F}_0$ can have negative values; see formula (2.10) on page 15 of Cox and Oakes (1984) for discrete distributions. Different from these existing works, our approach is based on the full likelihood function for (β_0, F_0) in Cox model (1), which, when profiling out the nuisance parameter F_0 by discretizing, is different from the likelihood function parameterized via (β_0, Λ_0). And our resulting MLE for F_0 is *always* a proper distribution function with values between 0 and 1, while our resulting MLE for β_0 show favorable simulation performances over Cox's estimator $\hat{\beta}_c$.

The main materials in this article are organized as follows. Section 2 derives the full likelihood function for (β_0, F_0) and 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 full-profile likelihood estimating function $\psi_n(\beta)$. 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 Sect. 4, we discuss a real dataset example, which demonstrates that in practical situation our full likelihood ratio test and Cox's partial likelihood ratio test may sometimes lead to statistically different conclusions. Some concluding remarks are included in Sect. 5.

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. 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 exact full likelihood of observed sample (2). In contrast, Bailey's estimator Bailey (1984) is based on an approximated likelihood function obtained by assuming a *discrete* baseline distribution F_0 (despite the fact that the Cox model assumes a continuous F_0 for continuous lifetime variables), thus his estimator is different from our actual MLE $\hat{\beta}_n$, and more importantly '*numerical studies suggest no small-sample advantage*' (Bailey 1984, page 734). 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. The result of Wilk's theorem is of interest here because it has been established generally for parametric full likelihoods, but not generally for nonparametric or semiparametric full likelihood functions, while the Cox model (1) is a semiparametric model. The last point (d) 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. However, the extension to other types of censored data requires additional and non-trivial work which will not be considered in this current paper.

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 the *full* likelihood function with restriction of all possible candidates for the MLE of F_0 to those distribution functions that assign all their probability masses to observations V_i's and interval $(V_{(n)}, \infty)$. In contrast, there have been some competing methods to estimate $F_0(t)$ 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 et al. (1993, Section IV.1.5) and Kalbfleisch and Prentice (2002, page 143).

2 Maximum likelihood estimators

For simplicity of presentation, this section considers the case that covariate Z is a scaler rather than a vector, i.e., $p = 1$ in (1). The generalization of our results to multivariate case with $p > 1$ is straightforward and is summarized at the end of this section.

To derive the full likelihood function for (β_0, F_0), we notice that under the assumption of Cox model (1), each X_i has a d.f. that satisfies the following Lehmann family properties (Kalbfleisch and Prentice 2002, page 97):

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

where $c_i = \exp(Z_i \beta)$ with $\beta = \beta_0$, $\bar{F}_0(t) = [1 - F_0(t)]$ 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 $F(t \mid Z_i)$ and $F_0(t)$, respectively. We also notice that if $g(v, \delta \mid z)$ represents the conditional p.d.f. of (V_i, δ_i) given $Z_i = z$ for survival data (2), we have $g(v, 1 \mid z) = f(v|z)\bar{F}_Y(v)$ and $g(v, 0 \mid z) = f_Y(v)\bar{F}(v|z)$. Thus, under the Cox model (1) with right censored survival data (2), the likelihood function of (V_i, δ_i) given $Z = Z_i$, $1 \le i \le n$, is given by

$$\prod_{i=1}^{n} g(V_i, \delta_i | Z_i) = \prod_{i=1}^{n} \left(f(V_i|Z_i)\bar{F}_Y(V_i)\right)^{\delta_i} \left(f_Y(V_i)\bar{F}(V_i|Z_i)\right)^{1-\delta_i},$$

which under (4) is proportional to

$$\prod_{i=1}^{n} (f(V_i \mid Z_i))^{\delta_i} \left(\bar{F}(V_i \mid Z_i)\right)^{1-\delta_i} = \prod_{i=1}^{n} \left(c_i f_0(V_i)\right)^{\delta_i} \left(\bar{F}_0(V_i)\right)^{c_i - \delta_i}.$$

Hence, the full likelihood function for (β_0, F_0) in Cox model (1) with data (2) is given by

$$L(\beta, F) = \prod_{i=1}^{n} (c_i \, dF(V_i))^{\delta_i} \left(\bar{F}(V_i)\right)^{c_i - \delta_i}, \tag{5}$$

where $c_i = \exp(Z_i \beta)$. To profile out nuisance parameter F_0, we, without loss of generality, assume that there are no ties among V_i's with $V_1 < \cdots < V_n$ (see Sect. 3 for treatment of ties), and we use the empirical likelihood parameterization (Owen 1988; Shorack and Wellner 1986). That is, we restrict all possible candidates for the MLE of F_0 to those d.f.'s that assign all their probability masses to points V_i's and interval (V_n, ∞), which writes the full likelihood function (5) as follows:

$$L(\beta, F) = \prod_{i=1}^{n} \left(c_i p_i\right)^{\delta_i} \left(\sum_{j=i+1}^{n+1} p_j\right)^{c_i - \delta_i}, \tag{6}$$

where $p_i = F(V_i) - F(V_i-)$ for $1 \le i \le n$; $0 \le p_{n+1} \le 1$ is the probability mass of d.f. F on interval (V_n, ∞); and we have $F(x) = \sum_{i=1}^{n} p_i I\{V_i \le x\}$ satisfying $\sum_{i=1}^{n+1} p_i = 1$.

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 $L(\beta, F)$ in (6) is maximized by:

$$1 - \hat{F}_n(t) = \prod_{V_i \le t} \frac{d_i - \delta_i}{d_i}. \tag{7}$$

In (6), we replace F by $\hat{F}_n$, then from the proof of (7) given in the Appendix (see Eq. (14)), we obtain the following *full-profile likelihood function* for β_0:

$$\ell(\beta) = \prod_{i=1}^{n} \left(\frac{c_i}{d_i}\right)^{\delta_i} \left(\frac{d_i - \delta_i}{d_i}\right)^{d_i - \delta_i} = \ell_c(\beta) \prod_{i=1}^{n} \left(\frac{d_i - \delta_i}{d_i}\right)^{d_i - \delta_i}, \tag{8}$$

where $\ell_c(\beta) = \prod_{i=1}^{n} (c_i/d_i)^{\delta_i}$ is the Cox's partial likelihood function. Thus, the MLE for β_0 is given by the solution $\hat{\beta}_n$ which maximizes the value of $\ell(\beta)$, and consequently $\hat{F}_n$ in (7) with β replaced by $\hat{\beta}_n$ is the MLE for F_0.

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

$$\psi_n(\beta) \equiv n^{-1} \frac{\mathrm{d}}{\mathrm{d}\beta} (\log \ell(\beta)) = n^{-1} \sum_{i=1}^{n} \delta_i \left(Z_i + e_i \log \frac{d_i - 1}{d_i} \right) = 0, \tag{9}$$

where $e_i = \sum_{j=i}^{n} Z_j c_j$, $\psi_n(\beta)$ is the *full-profile likelihood estimating function*, and from (8), $\log 0$ is set to 0 whenever it occurs. To compute $\hat{\beta}_n$, the Newton–Raphson method may be used with Cox's partial likelihood estimator $\hat{\beta}_c$ as the initial value.

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 that for fixed β: (a) Full likelihood function $L(\beta, F)$ in (6) has a finite maximum value; (b) $\hat{F}_n$ given by (7) maximizes $L(\beta, F)$ and is a proper d.f. with all terms on the right-hand side of equation between 0 and 1, because $d_i > c_n \ge 1$ for all $1 \le i < n$. These clearly are crucial reasons for having a well-defined full-profile likelihood function $\ell(\beta)$ as well as a well-defined MLE for F_0. 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 (6)-(9), we replace Z_i and c_i by $\tilde{Z}_i$ and $\tilde{c}_i = \exp(\tilde{Z}_i \beta)$, respectively, then we have $\tilde{c}_n \equiv 1$ for any β; the solution of (9) gives the MLE for β_0, still denoted as $\hat{\beta}_n$; and the resulting estimator in (7), still denoted by $\hat{\bar{F}}_n$, is the MLE for $[\bar{F}_0(t)]^{e^{Z_n \beta}}$, thus the MLE for $\bar{F}_0(t)$ is given by $[\hat{\bar{F}}_n(t)]^{e^{-Z_n \hat{\beta}_n}}$. Our extensive simulation studies show that such a treatment on condition $c_n \ge 1$ gives excellent performance on the resulting MLE $\hat{\beta}_n$ and the Newton–Raphson algorithm. In particular, $\tilde{c}_n = 1$ for any positive or negative β plays a crucial role for stable performance of the Newton–Raphson algorithm, because $\tilde{Z}_i$'s in Eq. (9) ensure $\tilde{d}_i > 1$ for all $1 \leq i < n$ and for any positive or negative β. Finally, we note 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 $\ell_c(\beta)$ in (8), thus does not affect $\hat{\beta}_c$, because $\ell_c(\beta) = \prod_{i=1}^n (c_i/d_i)^{\delta_i} = \prod_{i=1}^n (\tilde{c}_i/\tilde{d}_i)^{\delta_i}$ and the partial likelihood estimating function in (3) satisfies

$$\varphi_n(\beta) = n^{-1} \frac{\mathrm{d}}{\mathrm{d}\beta} (\log \ell_c(\beta)) = n^{-1} \sum_{i=1}^n \delta_i \left(Z_i - \frac{e_i}{d_i} \right) = n^{-1} \sum_{i=1}^n \delta_i \left(\tilde{Z}_i - \frac{\tilde{e}_i}{\tilde{d}_i} \right), \tag{10}$$

where $\tilde{d}_i = \tilde{c}_i + \cdots + \tilde{c}_n$ and $\tilde{e}_i = \tilde{Z}_i \tilde{c}_i + \cdots + \tilde{Z}_n \tilde{c}_n$. The uniqueness of the MLE $\hat{\beta}_n$ and further rationale for using the given $\tilde{Z}_i$ is discussed in Remark 2.

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:

$$\psi_n(\beta) = \varphi_n(\beta) + O_p\left(\frac{\log n}{n}\right), \tag{11}$$

where $\varphi_n(\beta)$ is the partial likelihood estimating function in (10). Further, Wilk's theorem on the log full-likelihood ratio is established below with proof given in the Appendix.

Theorem 1 *Assume* (11) *and assume the regularity conditions on Cox model* (1) (Andersen and Gill, 1982). *Then,* $R_0 = -2\log[\ell(\beta_0)/\ell(\hat{\beta}_n)]$ *converges in distribution to a chi-squared distribution with 1 degree of freedom as* $n \to \infty$.

Remark 2 Uniqueness of the MLE $\hat{\beta}_n$: To further understand the rationale of using $\tilde{Z}_i = Z_i - Z_n, 1 \leq i \leq n$, as suggested in Remark 1, we assume that the parameter space for β satisfies $|\beta| \leq M_1 < \infty$ for some constant M_1, and that the covariate variable Z has a finite support. Then, there exists a constant $0 < M < \infty$ such that we always have $|Z_i\beta| \leq M$ for $1 \leq i \leq n$ and any $|\beta| \leq M_1$, and we can rewrite Cox model (1) as $\lambda(t;\, z) = \lambda_M(t)e^{z\beta+M}$ for $\lambda_M(t) = e^{-M}\lambda_0(t)$, which gives $C_i = e^M c_i = e^{Z_i\beta+M} \geq 1$ for all $1 \leq i \leq n$ and for any $|\beta| \leq M_1$. With these C_i's, the arguments in (6)–(8) show that Eq. (9) becomes $\psi_{M,n}(\beta) \equiv n^{-1}\sum_{i=1}^n \delta_i \left(Z_i + E_i \log \frac{D_i - 1}{D_i}\right) = 0$ with $D_i = \sum_{j=i}^n C_j$ and $E_i = \sum_{j=i}^n Z_j C_j$, the resulting estimator $\hat{\bar{F}}_{M,n}$ by Eq. (7) is the MLE for $[\bar{F}_0(t)]^{e^{-M}}$, and Eq. (11) holds for $\psi_{M,n}(\beta)$ for fixed M as follows:

$$\psi_{M,n}(\beta) = \varphi_n(\beta) + O_p\left(\frac{\log n}{n}\right), \quad \text{as } n \to \infty \tag{12}$$

where $\varphi_n(\beta)$ is given by (10). The solution of equation $\psi_{M,n}(\beta) = 0$ gives an MLE $\hat{\beta}_{M,n}$ for β_0. Thus, the MLE for β_0 depends on the choice of M; in turn, it is not unique due to the Cox model assumption (1) itself, which allows the use of an arbitrarily large enough constant M in the way described above. On the other hand, by

Taylor's expansion and a proof similar to that for (11), we can show that for any given survival sample (2),

$$\psi_{M,n}(\beta) = \varphi_n(\beta) + o(1), \quad \text{as } M \to \infty \tag{13}$$

where $o(1)$ is uniformly for β. Equations (12)–(13) together explain why the Cox's partial likelihood estimator $\hat{\beta}_c$ has the well-known *good* asymptotic properties, despite it is not the *actual* MLE. With (13) in mind, in practice we should choose M considering the following two factors: (i) When $Z_n\beta \geq 0$, choose $M = 0$, i.e., if $Z_n\beta \geq 0$, we have $c_n \geq 1$, thus there is no need to use M; (ii) When $Z_n\beta < 0$, choose a small $M > 0$ to have $C_n \geq 1$, because a large $M > 0$ implies $\bar{\hat{F}}_{M,n}(t) \approx [\bar{F}_0(t)]^{e^{-M}} \approx 1$ and $\psi_{M,n}(\beta) \approx \varphi_n(\beta)$. Hence, in practice we use the smallest $M \geq 0$ under condition $C_n = e^{Z_n\beta+M} \geq 1$ for any $|\beta| \leq M_1$, and such an M is given by $M_n = \min\{M \geq 0 \,|\, Z_n\beta + M \geq 0\} = -Z_n\beta I\{Z_n\beta < 0\}$. Thus, for positive covariates Z_i's, we should use $M_n = 0$ if $\beta_0 > 0$; $M_n = -Z_n\beta$ if $\beta_0 < 0$. Since in practice the true value of β_0 is unknown, we always use $M_n = -Z_n\beta$ (unless there is evidence for $Z_n\beta_0 > 0$), which gives $Z_i\beta + M_n = \tilde{Z}_i\beta$ as suggested in Remark 1. In conclusion, the MLE for β_0 in the Cox model (1) is generally not uniquely defined; but *it is* uniquely defined under assumptions: (a) we know nothing about the true value of β_0; (b) $c_n = e^{Z_n\beta} \geq 1$ for any positive or negative β to ensure a well-defined full-profile likelihood function $\ell(\beta)$ in (8). And such unique MLE is given by the $\hat{\beta}_n$ in Remark 1. Note that under Remark 1, the MLE $\hat{\beta}_n$ is uniquely well defined as a solution of Eq. (9) with $\tilde{Z}_i$'s in the places of Z_i's, and Eq. (11) and Theorem 1 hold with $\tilde{Z}_i$'s in the places of Z_i's.

Remark 3 *p-Dimensional* Z_i: If Z_i and β_0 are p-dimensional vectors with $p > 1$ in (1)–(2), with minor modifications in the derivations and proofs we have that (6)–(11) and Remarks 1-2 hold, where with $c_i = \exp(Z_i^\top\beta)$, e_i and $\tilde{Z}_i = Z_i - Z_n$ are p-dimensional vectors, and (9) 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 degrees of freedom.

3 Simulations and treatment of ties

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. Routines in FORTRAN for computing $\hat{\beta}_n$ are available from the authors.

3.1 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 interval (0, 1). In our simulation studies, we consider $F_Y =$ Exp(2) as the d.f. of the right censoring variable Y_i, $F_Z =$ U(0, 1) as the d.f. of Z, and

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

Sample size	Parameter					
	$\beta_0 = 1$		$\beta_0 = 0$		$\beta_0 = -1$	
	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 2 Comparison between $\hat{\beta}_c$ and $\hat{\beta}_n$ ($\beta_0 = -2$, $n = 15$, censoring percentage = 57.0%)

Estimator	Simulation mean (SD)	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

$F_{X|Z} = \text{Exp}(e^{-Z\beta_0})$ 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 (SD) given in the parenthesis next to them, respectively. The censoring percentage in each case is also reported in Table 1.

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 $\beta_0 = \pm 2, \pm 3, \ldots$ 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 in Table 2, the ratio of simulation relative MSE for $\hat{\beta}_n$ and $\hat{\beta}_c$ is $8.969/33.658 = 26.6\%$. Also, note that with fixed censoring variable Exp(2), the censoring percentage increases as we choose a smaller and smaller β_0; thus our extensive simulation studies include cases with much higher censoring percentages than that shown in Table 2, and the results are similar to what's presented here.

Finally, although not presented here, our simulation studies also show that the MLE $\hat{F}_n$ given in (7) provides a very good estimate for $1 - [\bar{F}_0(t)]^{e^{Z_n \beta}}$; see Remark 1.

3.2 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 Therneau and Grambsch 2000, pages 48–49) d_i's in Cox's partial likelihood function $\ell_c(\beta)$ in (8) 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 $\hat{\beta}_E$. The idea of this Efron's estimation is that since the ties are caused by rounding errors in practice and we do not really know if it in fact is $V_1 < V_2$ or $V_2 < V_1$ (which affects the values of d_1 and d_2), we take '*average*' to modify d_i's in $\ell_c(\beta)$; see Therneau and Grambsch (2000). In our full-profile likelihood function $\ell(\beta)$ given in (8), we face the same problem with d_i's when there are ties among V_i's; thus Efron's estimation for d_i's is applicable in $\ell(\beta)$ for the same reason as for $\ell_c(\beta)$. Hence, applying these Efron's modified d_i's in (8)–(9), the MLE for β_0 when there are ties among V_i's is given by the solution of (9), still denoted by $\hat{\beta}_n$.

Some simulation results are presented in Tables 3 and 4 to compare the MLE $\hat{\beta}_n$ with $\hat{\beta}_E$. In these simulation studies, we consider $n = 15$, $F_Y = \text{Exp}(2)$, 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 $F_Z = \text{Exp}(1)$ for $\beta_0 = -0.75$ and $\beta_0 = 0.75$, respectively.

Overall, Tables 3 and 4 show that the MLE $\hat{\beta}_n$ performs favorably. In particular, note that in Table 3, the ratio of simulation relative MSE for $\hat{\beta}_n$ and $\hat{\beta}_E$ is

Table 3 Comparison between $\hat{\boldsymbol{\beta}}_E$ and $\hat{\boldsymbol{\beta}}_n$

$F_Z = \text{U}(0, 1)$	Estimator	Simul. mean (SD)	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 $\hat{\boldsymbol{\beta}}_E$ and $\hat{\boldsymbol{\beta}}_n$

$F_Z = \text{Exp}(1)$	Estimator	Simul. mean (SD)	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
β_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

Table 5 Stanford heart transplant data

	Data information		Estimator		p-Value		
Observations used	n	Censored obs.	$\hat{\beta}_c$	$\hat{\beta}_n$	Wald test	PLR test	FLR test
No. 76–100	25	8	0.367	0.397	0.063	0.056	0.038
No. 50–100	51	23	0.153	0.149	0.050	0.045	0.049

$7.785/20.711 = 37.59\%$ for $\beta_0 = -2$, while in Table 4, such ratio is $4.701/11.259 = 41.8\%$ for $\beta_0 = -0.75$.

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 that different likelihood ratio tests can lead to different conclusions in practice for smaller data sets, we use observations number 76–100 and observations number 50–100, respectively, to test $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$ 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 have statistically different conclusions for sample size $n = 25$ when 5% is used as the test significance level.

Of course, the point of this example is not to say that the FLR test is always more efficient than the PLR test; rather it demonstrates the fact that in practical situations FLR test and PLR test may sometimes lead to statistically different conclusions for small or moderate samples. This, along with the simulation results in Tables 1, 2, 3 and 4, indicates that the full-likelihood inferences for the Cox's model developed in this article are worth further examination and consideration.

5 Concluding remarks

We have derived the full likelihood function for (β_0, F_0) in the Cox model (1) with right censored survival data (2), and have derived joint MLE $(\hat{\beta}_n, \hat{F}_n)$ for (β_0, F_0), where $\hat{F}_n$ is always a proper distribution function. While our simulation studies indicate that for small or moderate sample sizes the MLE $\hat{\beta}_n$ performs favorably over Cox's partial likelihood estimator $\hat{\beta}_c$, the theoretical estimation bias of our MLE $\hat{\beta}_n$ in comparison with that of the Cox's estimator $\hat{\beta}_c$ is to be investigated in a forthcoming paper.

A preliminary version of the R-code that computes the multivariate p-dimensional MLE $\hat{\beta}_n$ and related quantities has been developed, and is now available at the following link: http://www.ms.uky.edu/~mai/Rcode/PHreg.fit.R.txt. Although the code needs to be further refined for publication quality, the computations we have conducted so far show that the algorithm is stable for the multivariate setting.

Acknowledgment The authors thank the Associate Editor and two referees for their comments and suggestions on the earlier draft of this article.

Appendix

Proof of (7) Let $a_i = p_i / b_i$ and $b_i = \sum_{j=i}^{n+1} p_j$. Then, we have

$$b_1 = 1, \quad b_{n+1} = p_{n+1}, \quad b_{i+1} = (b_i - p_i), \quad (1 - a_i) = b_{i+1}/b_i.$$

From $\prod_{i=1}^{n} (1 - a_i) = b_{n+1}$ and from

$$\begin{aligned}\prod_{i=1}^{n} (a_i)^{c_i} (1 - a_i)^{n-h_i} &= \left(\prod_{i=1}^{n} (a_i)^{c_i} \right) \prod_{i=1}^{n} \left(\frac{b_{i+1}}{b_i} \right)^{n-h_i} \\ &= \left(\prod_{i=1}^{n} (a_i)^{c_i} \right) (b_{n+1})^{n(1-\bar{c})} \prod_{i=1}^{n} (b_i)^{c_i},\end{aligned}$$

where $h_i = c_1 + \cdots + c_i$ and $\bar{c} = n^{-1} \sum_{i=1}^{n} c_i$, we can rewrite (6) as

$$\begin{aligned}L(\beta, F) &= \prod_{i=1}^{n} (c_i p_i)^{\delta_i} (b_i - p_i)^{c_i - \delta_i} = \left(\prod_{i=1}^{n} (c_i)^{\delta_i} (p_i)^{c_i} \right) \prod_{i=1}^{n} \left(\frac{1 - a_i}{a_i} \right)^{c_i - \delta_i} \\ &= \left(\prod_{i=1}^{n} (c_i)^{\delta_i} (p_i)^{c_i} \right) \frac{\prod_{i=1}^{n} (a_i)^{\delta_i} (1 - a_i)^{[n - \delta_i - (c_1 + \cdots + c_{i-1})]}}{\prod_{i=1}^{n} (a_i)^{c_i} (1 - a_i)^{[n - (c_1 + \cdots + c_i)]}} \\ &= \left(\prod_{i=1}^{n} (c_i)^{\delta_i} (p_i)^{c_i} \right) \frac{\prod_{i=1}^{n} (a_i)^{\delta_i} (1 - a_i)^{[n - \delta_i - (c_1 + \cdots + c_{i-1})]}}{(b_{n+1})^{n(1-\bar{c})} \prod_{i=1}^{n} (p_i)^{c_i}} \\ &= \prod_{i=1}^{n} (c_i a_i)^{\delta_i} (1 - a_i)^{d_i - \delta_i}. \end{aligned} \tag{14}$$

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 $\hat{a}_i = \delta_i / d_i$, $1 \le i \le n$, and it maximizes $L(\beta, F)$ under condition $c_n \ge 1$. Hence, (7) follows from noting that the $\hat{p}_i$'s corresponding to $\hat{a}_i$'s give $\bar{\hat{F}}_n(t) = \prod_{V_i \le t} (1 - \hat{a}_i)$ and that condition $c_n \ge 1$ implies all $0 \le \hat{a}_i \le 1$. □

Proof of (11) From Remark 1, we give the proof for $\psi_n(\beta)$ in (9) with $\tilde{Z}_i = Z_i - Z_n$, $1 \le i \le n$, assuming that $|\beta| \le M_\beta < \infty$ and Z has a finite support. From $\tilde{c}_n = 1$, $\tilde{Z}_n = 0$, Eq. (10) and Taylor's expansion, we have in (9),

$$\psi_n(\beta) = n^{-1} \sum_{i=1}^{n} \delta_i \tilde{Z}_i - n^{-1} \sum_{i=1}^{n-1} \delta_i \tilde{e}_i \left(\frac{1}{\tilde{d}_i} + \frac{1}{2\xi_i^2} \right) = \varphi_n(\beta) - \frac{1}{2} R_n, \tag{15}$$

where $R_n = n^{-1}\sum_{i=1}^{n-1}(\delta_i\tilde{e}_i)/\xi_i^2$ with ξ_i between $\tilde{d}_i$ and $(\tilde{d}_i - 1)$. From $(1 - 1/\tilde{d}_{n-1})^{-1} = (1 + 1/\tilde{c}_{n-1})$ and $m_i(\beta) = \min\{\tilde{Z}_j\beta \,|\, i \le j \le n\}$, the proof follows from

$$|R_n| \le \frac{1}{n}\sum_{i=1}^{n-1}\frac{|\tilde{e}_i|}{(\tilde{d}_i-1)^2} = \frac{1}{n}\sum_{i=1}^{n-1}\frac{|\tilde{e}_i|}{\tilde{d}_i^2(1-1/\tilde{d}_i)^2} \le \frac{1}{n}\sum_{i=1}^{n-1}\frac{|\tilde{e}_i|}{\tilde{d}_i^2}\frac{1}{(1-1/\tilde{d}_{n-1})^2}$$
$$\le O_p(n^{-1})\max_{1\le i\le n}|\tilde{Z}_i|\sum_{i=1}^{n-1}\frac{\exp(-m_i(\beta))}{\exp(\tilde{Z}_i\beta - m_i(\beta)) + \cdots + \exp(\tilde{Z}_n\beta - m_i(\beta))}$$
$$\le O_p(n^{-1})\sum_{i=1}^{n-1}\frac{1}{n-i+1} = O_p\left(\frac{\log n}{n}\right). \qquad (16)$$

□

Proof of Theorem 1 From Remark 1, we give the proof for $\ell(\beta)$ in (8) and $\psi_n(\beta)$ in (9) with $\tilde{Z}_i = Z_i - Z_n, 1 \le i \le n$, which implies $\tilde{c}_n = 1$ and $\tilde{Z}_n = 0$. Applying Taylor's expansion on $\log \ell(\beta_0)$ at point $\hat{\beta}_n$, we have that from (9) and $\psi_n(\hat{\beta}_n) = 0$,

$$R_0 = -n\psi_n'(\xi)(\beta_0 - \hat{\beta}_n)^2 = -\psi_n'(\xi)\left[\sqrt{n}(\hat{\beta}_n - \beta_0)\right]^2, \qquad (17)$$

where ξ is between $\hat{\beta}_n$ and β_0. From (11), we know that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\beta}_c - \beta_0)$ have the same limiting distribution $N(0, \sigma^2)$ 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, $\tilde{c}_n = 1$, $\tilde{Z}_n = 0$ and (10), we have that in (9)

$$-\psi_n'(\beta) = -n^{-1}\sum_{i=1}^{n-1}\delta_i\left(\frac{d\tilde{e}_i}{d\beta}\log\frac{\tilde{d}_i - 1}{\tilde{d}_i} + \frac{\tilde{e}_i^2}{\tilde{d}_i(\tilde{d}_i - 1)}\right)$$
$$= -n^{-1}\sum_{i=1}^{n-1}\delta_i\left\{\frac{d\tilde{e}_i}{d\beta}\left(-\frac{1}{\tilde{d}_i} - \frac{1}{2\xi_i^2}\right) + \frac{\tilde{e}_i^2}{\tilde{d}_i}\left(\frac{1}{\tilde{d}_i} + \frac{1}{\eta_i^2}\right)\right\}$$
$$= -\varphi_n'(\beta) + \frac{1}{2}R_{1,n} - R_{2,n}, \qquad (18)$$

where $R_{1,n} = n^{-1}\sum_{i=1}^{n-1}\delta_i(d\tilde{e}_i/d\beta)/\xi_i^2$ and $R_{2,n} = n^{-1}\sum_{i=1}^{n-1}\delta_i\tilde{e}_i^2/(\tilde{d}_i\eta_i^2)$ with ξ_i and η_i between $\tilde{d}_i$ and $(\tilde{d}_i - 1)$. Applying the argument in (16) 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). □

References

Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics, 10*, 1100–1120.

Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer.

Bailey, K. R. (1984). Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model. *The Annals of Statistics, 12*, 730–736.

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

Chang, M. N., Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics, 15*, 1536–1547

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

Cox, D. R., Oakes, D. (1984). *Analysis of survival data.* London: Chapman & Hall.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of American Statistical Association, 72*, 557–565.

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

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

Groeneboom, P., Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkhäuser Verlag.

Gu, M. G., Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics, 21*, 611–624.

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

Kalbfleisch, J. D., Prentice, R. L. (2002). *The statistical analysis of failure time data*. New York: Wiley.

Kay, R. (1979). Some further asymptotic efficiency calculations for survival data regression models. *Biometrika, 66*, 91–96.

Li, G., Li, R. Z., Zhou, M. (2005). Empirical likelihood in survival analysis. In J. Fan, G. Li (Eds.), *Contemporary multivariate analysis and design of experiments* (pp. 337–350). Singapore: The World Scientific Publisher.

Murphy, S. A., van der Vaart, A. W. (2000). On profile likelihood. *Journal of American Statistical Association, 95*, 449–465.

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.

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

Pan, X. R. (1997). *Empirical likelihood ratio for censored data*. Ph.D. Dissertation, Department of Statistics, University of Kentucky.

Shorack, G. R., Wellner, J. A. (1986). *Empirical processes with applications to statistics*. New York: Wiley.

Therneau, T. M., Grambsch, P. M. (2000). *Modeling survival data, extending the Cox model*. Berlin: Springer.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics, 9*, 93–108.