

## EMPIRICAL LIKELIHOOD FOR THE ADDITIVE RISK MODEL

BY

XUEWEN LU\* (CALGARY, ALBERTA)  
AND YONGCHENG QI (DULUTH, MINNESOTA)

*Abstract.* In this article, we investigate the empirical likelihood method for the additive risk model when the failure times are subject to left-truncation and right-censoring. An empirical likelihood ratio for the  $p$ -vector of regression coefficients is defined and it is shown that its limiting distribution is a weighted sum of independent chi-squared distributions with one degree of freedom. This enables one to make empirical likelihood based inference for the regression parameters. Finite sample performance of the proposed methods is illustrated in simulation studies to compare the empirical likelihood method with the normal-approximation-based method.

**2000 AMS Subject Classification:** 62E20, 62N01.

**Key words and phrases:** Additive risk model; empirical likelihood; left-truncation and right-censoring; normal approximation.

### 1. INTRODUCTION

In survival analysis, the multiplicative risk model (Cox [4]) and the additive model (Aalen [1]) provide the two principal frameworks for the regression analysis of censored survival data. The additive risk model provides a useful alternative to the multiplicative risk model. Buckley [3] pointed out that the additive risk model is biologically more plausible than the proportional hazards model, while O'Neill [7] found that the use of the proportional hazards model may result in serious bias when the true model is additive.

Let  $T$  be the survival time associated with a  $p$ -vector of possible time-varying covariates  $Z(t)$ . Then, under the additive risk model, the hazard function takes the form

$$(1) \quad \lambda(t; Z) = \lambda_0(t) + \beta_0^T Z(t),$$

---

\* Xuewen Lu's research was partially supported by NSERC Discovery Grant 73-1069 of Canada.

where  $\beta_0$  is a  $p$ -vector of regression parameters, and  $a^T$  is the transpose of a column-vector of  $a$ , and  $\lambda_0(t)$  is the baseline hazard function. In many applications such as biomedical and insurance problems, the survival time  $T$  is often subject to left-truncation and right-censoring, the baseline hazard function  $\lambda_0(t)$  is not parametrized. It is desirable to make statistical inference about the regression parameters  $\beta_0$  under the additive risk model. Lin and Ying [6] constructed a simple semiparametric estimating function for  $\beta_0$  and provided some semiparametric inference procedures. They also wrote a FORTRAN program to implement their methods for analyzing the additive risk model, so that the task of making inference becomes easy and reliable. An application of their methods can be found in Yip et al. [13].

We first outline the inference procedures by Lin and Ying [6]. Consider a set of  $n$  independent subjects such that the counting process  $\{N_i(t); t \geq 0\}$  for the  $i$ th subject in the set records the number of observed events up to time  $t$ . Assume  $Y_i(t)$  is a 0-1 predictable process indicating, by the value 1, whether the  $i$ th subject is at risk at time  $t$ ,  $Z_i(t)$  is the covariance process for the  $i$ th subject. In detail, assume the survival times or responses  $T_i$  are not completely observable due to left-truncation and right-censoring by the random variables  $L_i$  and  $C_i$ . Let  $\tilde{T}_i = \min(T_i, C_i)$  and  $\delta_i = I[T_i \leq C_i]$ . We use  $I[A]$  for the indicator function of a set  $A$ . Thus, the survival times  $T_i$  are right-censored by the censoring variables  $C_i$  and are left-truncated in the sense that  $(\tilde{T}_i, \delta_i, Z_i)$  can be observed only when  $\tilde{T}_i \geq L_i$ . The data, therefore, consist of  $n$  observations  $(\tilde{T}_i, L_i, \delta_i, Z_i)$  with  $\tilde{T}_i \geq L_i, i = 1, \dots, n$ . Then

$$Y_i(t) = I[\tilde{T}_i \geq t \geq L_i]$$

and

$$N_i(t) = I[L_i \leq T_i \leq \min(t, C_i)] = I[L_i \leq T_i \leq t, \delta_i = 1].$$

The intensity function for  $N_i(t)$  is given by

$$(2) \quad Y_i(t) d\Lambda(t; Z_i) = Y_i(t) \{d\Lambda_0(t) + \beta_0^T Z_i(t) dt\}$$

under the additive risk model (1), and the cumulative baseline hazard function is given by

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du.$$

The counting process  $N_i(t)$  can be uniquely decomposed so that, for every  $i$  and  $t$ ,

$$(3) \quad N_i(t) = M_i(t) + \int_0^t Y_i(u) d\Lambda(u; Z_i),$$

where  $M_i(t)$  is a local square-integrable martingale (Anderson and Gill [2]). The cumulative baseline hazard function  $\Lambda_0(t)$  is estimated by

$$(4) \quad \hat{\Lambda}_0(\beta, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u) \beta^T Z_i(u) du\}}{\sum_{i=1}^n Y_i(u)}$$

Lin and Ying [6] proposed to estimate  $\beta_0$  from the following estimating function:

$$U(\beta) = \sum_{i=1}^n \int_0^\tau Z_i(t) \{dN_i(t) - Y_i(t) d\hat{\Lambda}_0(\beta, t) - Y_i(t) \beta^T Z_i(t) dt\} \quad \text{for } 0 < \tau \leq \infty,$$

which is equivalent to

$$(5) \quad U(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} \{dN_i(t) - Y_i(t) \beta^T Z_i(t) dt\} \\ = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) - \left[ \sum_{i=1}^n \int_0^\tau Y_i(t) \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} d(t) \right] \beta,$$

where  $a^{\otimes 2} = aa^T$  and

$$\bar{Z}(t) = \frac{\sum_{j=1}^n Y_j(t) Z_j(t)}{\sum_{j=1}^n Y_j(t)}.$$

The resulting estimator takes the explicit form

$$(6) \quad \hat{\beta} = \left[ \sum_{i=1}^n \int_0^\tau Y_i(t) \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} d(t) \right]^{-1} \left[ \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dN_i(t) \right].$$

It is easy to see that

$$U(\beta_0) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t)\} dM_i(t),$$

which is a martingale integral. It then follows that the random vector (6) converges weakly to a  $p$ -variate normal with mean zero and with a covariance matrix  $V$ , i.e.

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{L}} N(0, V),$$

where  $V$  is the covariance matrix. Let

$$\alpha_r(t) = E\{Y_1(t) Z_1^r(t)\}, \quad r = 0, 1,$$

$$A = E \int_0^\tau Y_i(t) \left\{ Z_i(t) - \frac{\alpha_1(t)}{\alpha_0(t)} \right\}^{\otimes 2} d(t), \quad \Sigma = E \int_0^\tau \left\{ Z_i(t) - \frac{\alpha_1(t)}{\alpha_0(t)} \right\}^{\otimes 2} dN_1(t).$$

Then  $V = A^{-1} \Sigma A^{-1}$ , which can be consistently estimated by  $V_n = A_n^{-1} \Sigma_n A_n^{-1}$ , with

$$A_n = n^{-1} \sum_{i=1}^n \int_0^{\tau} Y_i(t) \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} d(t) \xrightarrow{P} A,$$

$$\Sigma_n = n^{-1} \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2} dN_i(t) \xrightarrow{P} \Sigma.$$

Therefore the large sample  $(1-\alpha)$ -level confidence region for  $\beta_0$  based on the above normal-approximation-based method is given by

$$(7) \quad R_{1\alpha} = \{\beta: n(\hat{\beta} - \beta)^T V_n^{-1} (\hat{\beta} - \beta) \leq \chi_p^2(\alpha)\},$$

where  $\chi_p^2(\alpha)$  is the  $(1-\alpha)$ -th quantile of the chi-squared distribution with  $p$  degrees of freedom.

In this paper, we will investigate how to apply empirical likelihood for the additive risk model with left-truncated and right-censored data. Empirical likelihood was introduced by Owen [8], [9] for a mean vector for i.i.d. observations, and has been extended to a wide range of applications. Though the empirical likelihood based method was first applied in survival analysis by Thomas and Grunkemeier [11], it is until recently that empirical likelihood for censored data analysis interests some authors, for example, Wang and Jing [12] studied functionals of survival distribution, Qin and Jing [10] investigated the Cox regression model. An appealing feature of the empirical likelihood approach is that it produces confidence regions whose shape and orientation are determined entirely by the data. It has many advantages over some classical and modern methods, such as the normal-approximation-based method and the bootstrap method. In particular, it does not impose prior constraints on region shape, it does not require the construction of a pivotal quantity and the region is range-preserving and transformation-respecting (Hall and La Scala [5]). Therefore, we are motivated to study the empirical likelihood method in constructing confidence regions for  $\beta_0$  in model (1). We will consider time-varying covariates and survival data subject to both left-truncation and right-censoring. Simulations show that the empirical likelihood method and the normal-approximation-based method work equally well.

The paper is organized as follows. In Section 2, we introduce the empirical likelihood method to the additive risk model (1). We define an empirical likelihood ratio statistic for the unknown parameter  $\beta_0$  and show that its limiting distribution is a weighted sum of independent chi-squared distributions with one degree of freedom. In Section 3, some simulation studies are given for both the empirical likelihood method and the normal-approximation-based method and comparisons will be made between these two methods. Finally, the proofs of our main results are given in Section 4.

## 2. METHODOLOGY AND MAIN RESULTS

Let us first give some motivations for our definition of the empirical likelihood for  $\beta$ . Consider the following testing problem:

$$H_0: \beta = \beta_0 \text{ vs. } H_1: \beta \neq \beta_0.$$

Under  $H_0$ ,

$$\begin{aligned} U(\beta) &= U(\beta_0) = \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} dM_i(t) \\ &= \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt\}. \end{aligned}$$

Since  $M_i(t)$ ,  $i = 1, \dots, n$ , are i.i.d. martingales,  $EU(\beta_0) = 0$ . Therefore, the problem of testing whether  $\beta_0$  is the true parameter of  $\beta$  is equivalent to testing whether  $EU(\beta_0) = 0$ . This can be done using Owen's empirical likelihood method (Owen [9]). Since  $A_0(t)$  and  $M_i(t)$  are unknown, we need to define an estimated empirical likelihood, evaluated at the true value  $\beta_0$ . Let  $p = (p_1, \dots, p_n)$  be a probability vector, i.e.,  $\sum_{i=1}^n p_i = 1$  and  $p_i \geq 0$  for all  $i$ . For  $1 \leq i \leq n$ , let

$$W_{ni} = \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt\}.$$

Then, we can define an estimated empirical likelihood, evaluated at the true value  $\beta_0$  of  $\beta$ , by

$$L(\beta_0) = \max \prod_{i=1}^n p_i$$

subject to

$$\sum_{i=1}^n p_i W_{ni} = 0 \quad \text{and} \quad \sum_{i=1}^n p_i = 1.$$

Then, by the method of Lagrange multipliers, we can easily get

$$p_i = n^{-1} \{1 + \lambda^T W_{ni}\}^{-1}, \quad i = 1, \dots, n,$$

where  $\lambda = (\lambda_1, \dots, \lambda_p)^T$  is the solution of

$$(8) \quad \frac{1}{n} \sum_{i=1}^n \frac{W_{ni}}{1 + \lambda^T W_{ni}} = 0.$$

Note that  $\prod_{i=1}^n p_i$ , subject to  $\sum_{i=1}^n p_i = 1$ , attains its maximum  $n^{-n}$  at  $p_i = n^{-1}$ . So we define the empirical likelihood ratio at  $\beta_0$  by

$$R(\beta_0) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n \{1 + \lambda^T W_{ni}\}^{-1}.$$

The corresponding empirical log-likelihood ratio can be defined as

$$(9) \quad l(\beta_0) = -2 \log R(\beta_0) = 2 \sum_{i=1}^n \log \{1 + \lambda^T W_{ni}\},$$

where  $\lambda$  is the solution of (8).

We need the following assumptions to establish our main theorem.

(A.1) Assume that  $0 < \tau < \infty$  and  $P(\bar{T} > \tau) > 0$ , which implies  $P(T > \tau) > 0$  and  $P(C > \tau) > 0$ .

(A.2)  $\sup_{t \in [0, \tau]} |\hat{\alpha}_1(t) - \alpha_1(t)| = O_p(1)$ , where  $\hat{\alpha}_1(t) = \sum_{j=1}^n Y_j(t) Z_j(t)/n$ .

(A.3) Assume that

$$\Sigma_0 = \lim_{n \rightarrow \infty} \left\{ n^{-1} \sum_{i=1}^n \left[ \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) d\Lambda_0(t) \right]^{\otimes 2} \right\}$$

exists. Matrices  $A$  and  $\Sigma$  are positively definite.

(A.4) Assume that, for all  $i$ ,

$$E \left| \int_0^{\tau} \{Z_i(t) - \alpha_1(t)/\alpha_0(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt\} \right|^2 < \infty,$$

$$E \left| \int_0^{\tau} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right|^2 < \infty,$$

and

$$E \left| \int_0^{\tau} \{\alpha_1(t) \alpha_0^{-1}(t)\} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right|^2 < \infty.$$

The assumption (A.1) is needed for some technical reasons. In Lin and Ying's paper,  $\tau$  was taken to be  $\infty$ . In practice, one can choose  $\tau = \max_i \{\bar{T}_i\}$ . When  $f_{ni}(t) = Z_i(t)/n$  are non-random functions of bounded variation in  $t$  with  $\int_0^{\infty} f_{ni}(t) \leq K < \infty$ ,  $K$  is independent of  $n$  and  $i$  and  $\sup_i \sum_{i=1}^n f_{ni}^2(t) = O(\log n^{-(1+\delta)})$  for some  $\delta > 0$ , then according to Lemma 5.1 of Zhou [14], (A.2) is true since  $\sup_{t \in [0, \tau]} |\hat{\alpha}_1(t) - \alpha_1(t)| = o_p(1)$ .

**THEOREM 1.** Suppose in model (1) that  $\lambda_0(\cdot)$  is continuous and  $(Z_i^T, T_i, C_i, L_i)$ 's are i.i.d. replicates, where the failure, censoring and left-truncation time variables  $T_i, C_i$  and  $L_i$  are conditionally independent, given  $Z_i$ . The variable  $Z_i = Z_i(t)$  is the covariate process for the  $i$ th subject and is time-varying (depending on time). In particular,  $(Z_i^T, \bar{T}_i = \min(T_i, C_i), \delta_i = I[T_i \leq C_i], L_i)$ 's with  $\bar{T}_i \geq L_i$  are the observed data, where  $\delta_i$  is the censoring indicator. If  $\beta_0$  is the true value of  $\beta$ , then under the assumptions (A.1)–(A.4) the limiting distribution of  $l(\beta_0)$  is a weighted sum of independent chi-squared distributions with one degree of freedom. That is,

$$l(\beta_0) \xrightarrow{\mathcal{L}} w_1 \chi_{1,1}^2 + w_2 \chi_{1,2}^2 + \dots + w_p \chi_{1,p}^2,$$

where  $w_1, \dots, w_p$  are the eigenvalues of  $\Sigma_1^{-1} \Sigma$  for  $\Sigma_1 = \Sigma + \Sigma_0$  and  $\chi_{1,1}^2, \dots, \chi_{1,p}^2$  are independent standard chi-squared random variables with one degree of freedom.

We note that  $\Sigma_0$  can be consistently estimated by

$$\Sigma_{0n} = n^{-1} \sum_{i=1}^n \left[ \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) d\hat{\Lambda}_0(\beta, t) \right]^{\otimes 2},$$

$\Sigma$  and  $\Sigma_1$  can be consistently estimated by  $\Sigma_n$  and  $\Sigma_{1n} = \Sigma_n + \Sigma_{0n}$ , respectively. Then  $w_i, 1 \leq i \leq p$ , can be consistently estimated by the eigenvalues  $\hat{w}_i$ 's of  $\Sigma_{1n}^{-1} \Sigma_n$ .

A simple approach to construct a  $(1 - \alpha)$ -level confidence region for  $\beta$ , based on Theorem 1, is

$$(10) \quad R_{2\alpha} = \{ \beta : l(\beta) \leq \chi_p^2(\alpha) \}.$$

$R_{2\alpha}$  gives an approximate confidence region for  $\beta$  with asymptotically correct coverage probability  $1 - \alpha$ , i.e.,

$$P(\beta_0 \in R_{2\alpha}(\beta)) = (1 - \alpha) + o(1).$$

### 3. SOME SIMULATION STUDIES

A series of simulation studies will be carried out to compare the performance in terms of coverage probabilities between the empirical likelihood method and the normal-approximation-based method. To do this, we consider the following additive risk model with two covariates  $(Z_1, Z_2)$ ,  $\lambda_0(t) = 1$ , and

$$\lambda(t; Z_1, Z_2) = 1 + (0.3Z_1 - 0.02Z_2).$$

This model is adopted from the simulation studies for continuous-time recapture experiments in Yip et al. [13]. Originally, the survival times are the capture times. The variable  $Z_1$  corresponds to the sex of the animals, with half of the subjects assigned to each sex ( $Z_1 = 1$  corresponds to male,  $Z_1 = 0$  corresponds to female), and  $Z_2$  corresponds to weight, with a normal distribution of mean eight and variance four. This model implies that males are more catchable than females and the catchability declines with weight. In our current simulation studies, the censoring times  $C_i$ 's are generated from the uniform distribution  $U[0, c]$ , where  $c$ 's are chosen for a prespecified proportion of censoring. When  $c = 5, 3, 2$ , the corresponding censoring proportions are 20%, 32% and 44%, respectively. The left-truncation times  $L$ 's are fixed at constants 0.0 and 0.1, they are chosen for 0% and 12% proportion of truncation, respectively. The original sample size has been chosen to be 25, 50, and 100, respectively. The final simulated observations from the above additive risk model are  $(\tilde{T}_i, \delta_i, Z_{1i}, Z_{2i}, L_i)$ , with  $\tilde{T}_i \geq L_i, i = 1, \dots, n$ , where

$$\tilde{T}_i = \min(T_i, C_i), \quad \delta_i = I[X_i \leq C_i].$$

There are  $B = 1000$  simulation samples for each combination of  $n$ ,  $c$  and  $L$ . For the nominal confidence level  $1 - \alpha$  taken to be 0.95, Table 1 presents the results of approximation coverage probabilities for the empirical likelihood method and the normal-approximation-based method computed from the inequalities (7) and (10) based on these 1000 simulation samples.

TABLE 1. Coverage probabilities for  $\beta^T = (\beta_1, \beta_2)$  at level 0.95

Censoring proportion	$n$	Truncation $L = 0$		Truncation $L = 0.1$	
		Normal approximation	Empirical likelihood	Normal approximation	Empirical likelihood
20%	25	0.959	0.912	0.964	0.906
	50	0.960	0.948	0.951	0.936
	100	0.958	0.953	0.950	0.949
32%	25	0.965	0.916	0.966	0.890
	50	0.959	0.938	0.962	0.935
	100	0.960	0.957	0.960	0.955
44%	25	0.970	0.902	0.968	0.868
	50	0.968	0.943	0.953	0.927
	100	0.960	0.952	0.951	0.943

From these simulation studies, we can make the following conclusions. For the same truncation value and different censoring proportions, both the empirical likelihood and the normal-approximation-based methods have similar performances. When data are truncated, the performances of both methods are slightly worsened. At each nominal level, the coverage probabilities of both methods increase as the sample sizes increase. They appear to be close to the nominal levels with sample sizes  $n \geq 100$ . When sample sizes are small ( $n = 25$ ), the normal-approximation-based method tends to work better than the empirical likelihood method. But the coverage probabilities of both methods deviate from the nominal levels.

#### 4. PROOFS

To prove Theorem 1, we need the following lemmas.

LEMMA 1. Under the same conditions as in Theorem 1, we have

$$n^{-1/2} \sum_{i=1}^n W_{ni} = n^{-1/2} U(\beta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma).$$

LEMMA 2. Under the same conditions as in Theorem 1, we have

- (i)  $n^{-1} \sum_{i=1}^n W_{ni} W_{ni}^T \xrightarrow{P} \Sigma_1$ ,
- (ii)  $\max_{1 \leq i \leq n} |W_{ni}| = o_p(n^{1/2})$ .



Proof of Lemma 1. Write  $W_{ni}$  as

$$\begin{aligned}
 (11) \quad W_{ni} &= \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt - Y_i(t) d\Lambda_0(t)\} \\
 &\quad + \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) d\Lambda_0(t) \\
 &= \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) dM_i(t) + \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) d\Lambda_0(t) \\
 &\triangleq P_{ni} + Q_{ni}.
 \end{aligned}$$

Hence, we have

$$\begin{aligned}
 n^{-1/2} \sum_{i=1}^n W_{ni} &= n^{-1/2} U(\beta_0) - n^{-1/2} \sum_{i=1}^n \int_0^{\tau} \{Z_i(t) - \bar{Z}(t)\} Y_i(t) d\Lambda_0(t) \\
 &= n^{-1/2} U(\beta_0) + o_p(1).
 \end{aligned}$$

The proof of Lemma 1 is completed by the standard counting process arguments about the random vector  $n^{-1/2} U(\beta_0)$ .

Proof of Lemma 2. From (11) we have

$$W_{ni} W_{ni}^T = P_{ni} P_{ni}^T + Q_{ni} Q_{ni}^T + P_{ni} Q_{ni}^T + Q_{ni} P_{ni}^T.$$

Since  $P_{ni}$  is a martingale, it is easy to see that, for any  $a \in \mathbb{R}^p$ ,

$$E \left\{ n^{-1} \sum_{i=1}^n a^T P_{ni} Q_{ni}^T a \right\}^2 = o(1) \quad \text{and} \quad E \left\{ n^{-1} \sum_{i=1}^n a^T Q_{ni} P_{ni}^T a \right\}^2 = o(1);$$

hence

$$n^{-1} \sum_{i=1}^n \{P_{ni} Q_{ni}^T + Q_{ni} P_{ni}^T\} = o_p(1).$$

Therefore, we have

$$\begin{aligned}
 n^{-1} \sum_{i=1}^n W_{ni} W_{ni}^T &= n^{-1} \sum_{i=1}^n P_{ni} P_{ni}^T + n^{-1} \sum_{i=1}^n Q_{ni} Q_{ni}^T + o_p(1) \\
 &= \Sigma + \Sigma_0 + o_p(1) = \Sigma_1 + o_p(1).
 \end{aligned}$$

This completes the proof of Lemma 2 (i).

To prove Lemma 2 (ii), let  $g(t) = \alpha_1(t)/\alpha_0(t)$ . We write  $W_{ni}$  as

$$(12) \quad W_{ni} = R_{ni} - S_{ni},$$

where

$$R_{ni} = \int_0^{\tau} \{Z_i(t) - g(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt\}$$

and

$$S_{ni} = \int_0^{\tau} \{Z_i(t) - g(t)\} \{dN_i(t) - Y_i(t) \beta_0^T Z_i(t) dt\}.$$

Let  $\hat{\alpha}_0(t) = \sum_{j=1}^n Y_j(t)/n$ ,  $\hat{\alpha}_1(t) = \sum_{j=1}^n Y_j(t) Z_j(t)/n$ ; then  $\bar{Z}(t) = \hat{\alpha}_1(t)/\hat{\alpha}_0(t)$ . By Lemma 5.1 of Zhou [14] and the assumption (A.1), we obtain

$$\sup_{t \in [0, \tau]} |\hat{\alpha}_0^{-1}(t)| = O_p(1) \quad \text{and} \quad \sup_{t \in [0, \tau]} |\{\hat{\alpha}_0(t) - \alpha_0(t)\}| = o_p(1).$$

Therefore, using the assumption (A.2) and the identity

$$\hat{\alpha}_1(t)/\hat{\alpha}_0(t) - \alpha_1(t)/\alpha_0(t) = \{\hat{\alpha}_1(t) - \alpha_1(t)\} \hat{\alpha}_0^{-1}(t) + \{\hat{\alpha}_0(t) - \alpha_0(t)\} \alpha_1(t) \alpha_0^{-1}(t) \hat{\alpha}_0^{-1}(t),$$

we have

$$\begin{aligned} & |\hat{\alpha}_1(t)/\hat{\alpha}_0(t) - \alpha_1(t)/\alpha_0(t)| \\ & \leq \sup_{t \in [0, \tau]} |\{\hat{\alpha}_1(t) - \alpha_1(t)\}| \sup_{t \in [0, \tau]} |\hat{\alpha}_0^{-1}(t)| + \sup_{t \in [0, \tau]} |\{\hat{\alpha}_0(t) - \alpha_0(t)\}| \\ & \quad \times \sup_{t \in [0, \tau]} |\hat{\alpha}_0^{-1}(t)| |\alpha_1(t) \alpha_0^{-1}(t)| \\ & = O_p(1) + o_p(1) |\alpha_1(t) \alpha_0^{-1}(t)|. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} |S_{ni}| & \leq O_p(1) \left| \int_0^{\tau} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right| \\ & \quad + o_p(1) \left| \int_0^{\tau} \{\alpha_1(t) \alpha_0^{-1}(t)\} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right|. \end{aligned}$$

Note that, by the assumption (A.4), and

$$E |R_{ni}|^2 < \infty, \quad E \left| \int_0^{\tau} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right|^2 < \infty,$$

$$E \left| \int_0^{\tau} \{\alpha_1(t) \alpha_0^{-1}(t)\} \{dN_i(t) + Y_i(t) \beta_0^T Z_i(t) dt\} \right|^2 < \infty,$$

following the proof of Lemma 3 in Owen [9], we obtain  $\max_i |R_{ni}| = o_p(n^{1/2})$  and  $\max_i |S_{ni}| = o_p(n^{1/2})$ . Therefore  $\max_i |W_{ni}| = o_p(n^{1/2})$ . The proof of Lemma 2 (ii) is completed.

**Proof of Theorem 1.** Write  $\lambda = \rho\theta$ , where  $\rho \geq 0$  and  $|\theta| = 1$ . Let  $\hat{S}_n = n^{-1} \sum_{i=1}^n W_{ni} W_{ni}^T$ . From the proof of Lemma 2 (i) we obtain

$$\theta^T \hat{S}_n \theta = \theta^T \Sigma_1 \theta + o_p(1).$$

Then, by Lemma 1 and Lemma 2 (i), and an argument similar to (2.14) in Owen [9], we can show that

$$(13) \quad |\lambda| = o_p(n^{-1/2}).$$

From (8) we have

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{W_{ni}}{1 + \lambda^T W_{ni}} = \frac{1}{n} \sum_{i=1}^n W_{ni} \left[ 1 - \lambda^T W_{ni} + \frac{(\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}} \right] \\ &= \frac{1}{n} \sum_{i=1}^n W_{ni} - \left( \frac{1}{n} \sum_{i=1}^n W_{ni} W_{ni}^T \right) \lambda + \frac{1}{n} \sum_{i=1}^n \frac{W_{ni} (\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}}. \end{aligned}$$

Hence

$$(14) \quad \lambda = \left( \sum_{i=1}^n W_{ni} W_{ni}^T \right)^{-1} \sum_{i=1}^n W_{ni} + \gamma_n,$$

where

$$\gamma_n = \left( n^{-1} \sum_{i=1}^n W_{ni} W_{ni}^T \right)^{-1} \frac{1}{n} \sum_{i=1}^n \frac{W_{ni} (\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}}.$$

Now Lemma 2 and (13) together prove

$$(15) \quad \gamma_n \leq O_p(n^{-1}) \max_{1 \leq i \leq n} |W_{ni}| = o_p(n^{-1/2})$$

by the fact that

$$\frac{1}{n} \sum_{i=1}^n \frac{W_{ni} (\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}} \leq 2|\lambda|^2 \left( \max_{1 \leq i \leq n} |W_{ni}| \right) n^{-1} \sum_{i=1}^n |W_{ni}|^2$$

in probability. Again by (8), we obtain

$$(16) \quad 0 = \sum_{i=1}^n \frac{\lambda^T W_{ni}}{1 + \lambda^T W_{ni}} = \sum_{i=1}^n (\lambda^T W_{ni}) - \sum_{i=1}^n (\lambda^T W_{ni})^2 + \sum_{i=1}^n \frac{(\lambda^T W_{ni})^3}{1 + \lambda^T W_{ni}}.$$

By Lemma 2 and (13), we have

$$(17) \quad \sum_{i=1}^n \frac{(\lambda^T W_{ni})^3}{1 + \lambda^T W_{ni}} \leq 2|\lambda|^3 \max_{1 \leq i \leq n} |W_{ni}| \sum_{i=1}^n |W_{ni}|^2 = o_p(1).$$

From (16) and (17), we get

$$(18) \quad \sum_{i=1}^n (\lambda^T W_{ni}) = \sum_{i=1}^n (\lambda^T W_{ni})^2 + o_p(1).$$

Applying Taylor's expansion to (9), using (14), (18) and Lemma 2, we obtain

$$\begin{aligned}
 (19) \quad l(\beta_0) &= 2 \sum_{i=1}^n \log \{1 + \lambda^T W_{ni}\} = 2 \sum_{i=1}^n \left\{ \lambda^T W_{ni} - \frac{1}{2} (\lambda^T W_{ni})^2 \right\} + r_n \\
 &= \sum_{i=1}^n \lambda^T W_{ni} W_{ni}^T \lambda + r_n \\
 &= (n^{-1/2} \sum_{i=1}^n W_{ni})^T (n^{-1} \sum_{i=1}^n W_{ni} W_{ni}^T)^{-1} (n^{-1/2} \sum_{i=1}^n W_{ni}) + r_n \\
 &= (\Sigma^{-1/2} n^{-1/2} \sum_{i=1}^n W_{ni})^T (\Sigma^{1/2} \Sigma_1^{-1} \Sigma^{1/2}) (\Sigma^{-1/2} n^{-1/2} \sum_{i=1}^n W_{ni}) + o_p(1) + r_n,
 \end{aligned}$$

where, for some constant  $C > 0$ ,

$$|r_n| \leq C \sum_{i=1}^n (\lambda^T W_{ni})^3 \leq C |\lambda|^3 \left( \max_{1 \leq i \leq n} |W_{ni}| \right) \sum_{i=1}^n |W_{ni}|^2 = o_p(1),$$

by the similar arguments to (17). Therefore, Theorem 1 follows straightaway from Lemma 1 using the fact that  $\Sigma^{1/2} \Sigma_1^{-1} \Sigma^{1/2}$  and  $\Sigma_1^{-1} \Sigma$  have the same eigenvalues.

#### REFERENCES

- [1] O. O. Aalen, *A model for nonparametric regression analysis of counting processes*, Lecture Notes in Statist. 2, N. Kłonecki, A. Kosek and J. Rosiński (Eds.), Springer, New York 1980, pp. 1–25.
- [2] P. K. Andersen and R. D. Gill, *Cox's regression model for counting processes: a large sample study*, Ann. Statist. 10 (1982), pp. 1100–1120.
- [3] J. D. Buckley, *Additive and multiplicative models for survival rates*, Biometrics 40 (1984), pp. 51–62.
- [4] D. R. Cox, *Regression models and life-tables (with discussion)*, J. Roy. Statist. Soc. Ser. B 34 (1972), pp. 187–220.
- [5] P. Hall and B. La Scala, *Methodology and algorithms of empirical likelihood*, Internat. Statist. Rev. 58 (1990), pp. 109–127.
- [6] D. Y. Lin and Z. Ying, *Semiparametric analysis of the additive risk model*, Biometrika 81 (1994), pp. 61–71.
- [7] T. J. O'Neill, *Inconsistency of the misspecified proportional hazards model*, Statist. Probab. Lett. 4 (1986), pp. 219–222.
- [8] A. Owen, *Empirical likelihood ratio confidence intervals for single functional*, Biometrika 75 (1988), pp. 237–249.
- [9] A. Owen, *Empirical likelihood ratio confidence regions*, Ann. Statist. 18 (1990), pp. 90–120.
- [10] G. Qin and B. Y. Jing, *Empirical likelihood for Cox regression model under random censorship*, Comm. Statist. Simulation 30 (2001), pp. 79–90.

- [11] D. R. Thomas and G. L. Grunkemeier, *Confidence interval estimation of survival probabilities for censored data*, J. Amer. Statist. Assoc. 70 (1975), pp. 866–871.
- [12] Q. Y. Wang and B. Y. Jing, *Empirical likelihood for a class of functionals of survival distribution with censored data*, Ann. Inst. Statist. Math. 53 (2001), pp. 517–527.
- [13] P. S. F. Yip, Y. Zhou, D. Y. Lin and X. Z. Fang, *Estimation of population size based on additive hazards models for continuous-time recapture experiments*, Biometrics 55 (1999), pp. 904–908.
- [14] M. Zhou, *Asymptotic normality of the ‘synthetic data’ regression estimator for censored survival data*, Ann. Statist. 20 (1992), pp. 1002–1021.

Xuwen Lu  
Department of Mathematics & Statistics  
University of Calgary  
Calgary, Alberta, T2N 1N4, Canada  
E-mail: lux@math.ucalgary.ca

Yongcheng Qi  
Department of Mathematics & Statistics  
University of Minnesota–Duluth  
Duluth, MN 55812, U.S.A.  
E-mail: yqi@d.umn.edu

*Received on 25.6.2004;*  
*revised version on 27.11.2004*

---

