Statistica Sinica 3(1993), 543-555

# ON THE BARTLETT ADJUSTMENT FOR THE PARTIAL LIKELIHOOD RATIO TEST IN THE COX REGRESSION MODEL

Minggao Gu and Zukang Zheng

*McGill University and Fudan University*

*Abstract:* The Bartlett adjustment for the partial likelihood ratio test in Cox regression model is established under one-dimensional parameter. If the baseline hazard is unspecified, the adjustment factor can be estimated from the data. The procedure give more accurate probability than the normal approximation to the log-rank test.

*Key words and phrases:* Adjustment, asymptotic expansion, partial likelihood, Cox regression model, survival data.

## 1. Introduction and Result

In parametric inference, the likelihood ratio test is one of the most popular statistic for inference. One reason for its popularity is the Bartlett (1937) adjustment to the likelihood ratio statistics. When the sample size is small, this adjustment may have a significant improvement over the ordinary asymptotic theory. For a detailed account and the proof, see Barndoff-Nielsen and Cox (1984).

In biomedical statistical inference, it is often unrealistic to make full parametric assumptions. In 1972, Cox introduced the partial likelihood approach which does not need full specification of the underlying distribution. As a special case, the Cox regression model has become the most widely used statistical tool for analyzing censored failure time data.

Early works on the Cox regression model have show that the partial likelihood score statistics are asymptotically normal, for example Tsiatis (1981), Andersen and Gill (1982), Naes (1982), Bailey (1983). As an easy consequence, one can deduce that the partial likelihood ratio statistics are asymptotically $\chi^2$. One problem that has been mentioned but remains unsolved (Barndoff-Nielsen and Cox (1984)) is whether the Bartlett adjustment or something similar works for partial likelihood. This paper gives a positive answer to the problem under the Cox regression model. The proof so far works for an one dimensional covariate but we believe that it can be extended to a multi-dimensional covariate.

Let $x_1^0, \ldots, x_n^0$ be a sequence of positive random variables, usually representing survival times of $n$ patients in a clinical trial. Let $z_1, \ldots, z_n$ be their corresponding covariates. The Cox (1972, 1975) regression model, for given $z$, assumes that the hazard function of $x^0$ satisfies

$$\lambda(t|z) = \lambda_0(t)\exp(\beta_0 z), \tag{1.1}$$

where $\lambda_0$ is called the baseline hazard function. In survival analysis, the $x^0$'s are usually not directly observable. We observe $(x_i, \delta_i)$, $i = 1, \ldots, n$, where $x_i = \min(x_i^0, c_i)$, $\delta_i = 1_{[x_i = x_i^0]}$ and $c_i$ is the censoring variable. In this paper, we assume that $x_i^0, c_i$ are independent given $z_i$, the hazard function of $c_i$ given $z$ does not depend on $\beta_0$, $z_i$ is uniformly bounded, and $y_i = (z_i, x_i^0, c_i)$ are i.i.d. The partial likelihood function for inference of $\beta_0$ is

$$L_n(\beta) = \prod_{i=1,\ldots,n; x_i \le T} \left[\frac{\exp(\beta z_i)}{\sum_{j=1}^n \exp(\beta z_j) 1_{[x_j \ge x_i]}}\right]^{\delta_i}, \tag{1.2}$$

where $T$ is such that $\Pr\{x_i \ge T\} > 0$. Let $\hat\beta$ be the Cox regression estimate, or be a number such that $L_n(\hat\beta) = \sup_\beta L_n(\beta)$. The partial likelihood ratio statistic is defined by

$$w = 2\{l_n(\hat\beta) - l_n(\beta_0)\}, \tag{1.3}$$

where $l_n(\beta) = \log L_n(\beta)$.

Before we introduce the main theorem, note that the Bartlett adjustment for the likelihood ratio statistics $w$ can be written as (one parameter case)

$$\Pr\{w \le u\} = \left(1 - \frac{1}{2}bn^{-1}\right)\chi_1^2(u) + \frac{1}{2}bn^{-1}\chi_3^2(u) + o(n^{-1}), \tag{1.4}$$

where $b$ is a constant.

The main theorem of this paper is the following.

**Theorem 1.1.** *Under the above assumptions,*

$$\begin{aligned}\Pr\{w \le u\} &= \chi_1^2(u) + \frac{1}{n}\left\{\left(\frac{\kappa_4}{8} - \frac{F}{2} - \frac{A^2}{2}\right)\chi_1^2(u) + \left(-\frac{\kappa_4}{4} + \frac{F}{2} + \frac{A^2}{2}\right)\chi_3^2(u)\right. \\ &\quad \left. + \frac{\kappa_4}{8}\chi_5^2(u)\right\} + O(n^{-3/2}),\end{aligned} \tag{1.5}$$

*where $\chi_p^2(u)$ is the value of the $\chi^2$ distribution with $p$ degree of freedom at $u$ and*

$$A = \frac{\Delta - 3\zeta}{6\sigma^3},$$

$$\begin{aligned} F &= \frac{3}{4} + \frac{1}{\sigma^6}\Big(\frac{7}{18}\Delta^2 - \frac{11}{6}\Delta\zeta + \frac{1}{4}\zeta^2\Big) \\ &\quad + \frac{1}{\sigma^4}\Big( - \frac{1}{4}\mu_2 + \mu_3 + 3\mu_4 + \frac{1}{2}\mu_5 - \mu_6 - \frac{1}{2}\mu_7 + 2\mu_8 - 2\mu_9\Big), \end{aligned} \tag{1.6}$$

$$\kappa_4 = -\, 9\frac{\zeta^2}{\sigma^6}. \tag{1.7}$$

*The constants involved on the right of* (1.6) *and* (1.7) *are defined below. For* $k = 0, \dots, 4$, $i = 1, 2, 3$

$$\begin{aligned} \alpha_k^{(i)}(t) &= E\Big[z^k e^{(i+1)\beta_0 z} 1_{[x\geq t]}\Big], \qquad \alpha_k(t) = \alpha_k^{(0)}(t), \\ \Lambda_k(t) &= \int_0^t \frac{\alpha_k(s)}{\alpha_0(s)}\lambda_0(s)ds, \qquad \Lambda_1^{(2)}(t) = \int_0^t \frac{\alpha_1^2(s)}{\alpha_0^2(s)}\lambda_0(s)ds, \\ \eta_k(t) &= \frac{\alpha_k^{(1)}(t)}{\alpha_0(t)}\Lambda_0(t) - \frac{\alpha_{k-1}^{(1)}(t)}{\alpha_0(t)}\Lambda_1(t) \\ \eta(t) &= -\Big(\eta_2(t) - \eta_1(t)\frac{\alpha_1(t)}{\alpha_0(t)}\Big), \\ \theta_k^{(i)}(t) &= \frac{\alpha_k^{(i)}(t)}{\alpha_0(t)} - 2\alpha_{k-1}^{(i)}(t)\frac{\alpha_1(t)}{\alpha_0^2(t)} + \alpha_{k-2}^{(i)}(t)\frac{\alpha_1^2(t)}{\alpha_0^3(t)}, \quad \theta_k(t) = \theta_k^{(0)}(t), \end{aligned} \tag{1.8}$$

*and*

$$\begin{aligned} \sigma^2 &= \int_0^T \theta_2(t)\alpha_0(t)\lambda_0(t)dt, \quad \Delta = -\int_0^T \Big(\theta_3(t) - \theta_2(t)\frac{\alpha_1(t)}{\alpha_0(t)}\Big)\alpha_0(t)\lambda_0(t)dt, \\ \zeta &= -\int_0^T \Big(\theta_3^{(1)}(t)\Lambda_0(t) - \theta_2^{(1)}(t)\Lambda_1(t)\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_2 &= \int_0^T \Big(\theta_4(t) - 2\theta_3(t)\frac{\alpha_1(t)}{\alpha_0(t)} + \theta_2(t)\frac{\alpha_1^2(t)}{\alpha_0^2(t)}\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_3 &= \int_0^T \Big(\eta_4(t) - 3\frac{\alpha_1(t)}{\alpha_0(t)}\eta_3(t) + 3\frac{\alpha_1^2(t)}{\alpha_0^2(t)}\eta_2(t) - \frac{\alpha_1^3(t)}{\alpha_0^3(t)}\eta_1(t)\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_4 &= \int_0^T \theta_2(t)\eta(t)\alpha_0(t)\lambda_0(t)dt, \\ \mu_5 &= \int_0^T \theta_2^{(1)}(t)\Big(\Lambda_2(t) - \Lambda_1^{(2)}(t)\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_6 &= \int_0^T \Big(\theta_4^{(2)}(t)\Lambda_0^2(t) - 2\theta_3^{(2)}(t)\Lambda_1(t)\Lambda_0(t) + \theta_2^{(2)}(t)\Lambda_1^2(t)\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_7 &= \int_0^T \Big(\theta_4^{(1)}(t)\Lambda_0(t) - 2\theta_3^{(1)}(t)\Lambda_1(t) + \theta_2^{(1)}(t)\Lambda_1^{(2)}(t)\Big)\alpha_0(t)\lambda_0(t)dt, \\ \mu_8 &= \int_0^T \eta^2(t)\alpha_0(t)\lambda_0(t)dt, \end{aligned} \tag{1.9}$$

$$\mu_9 = \int_0^T \theta_2^{(1)}(t) \left( \int_0^t \eta(s)\lambda_0(s)ds \right) \alpha_0(t)\lambda_0(t)dt.$$

**Remark 1.1.** The constants $A$, $F$ and $\kappa_4$ appearing in the adjustment (1.5) have statistical meanings. From the proof in the next section, we see that $w = X^2$, and $E[X] = A/\sqrt{n} + O(n^{-3/2})$, $\mathrm{Var}(X) = 1 + F/n + O(n^{-2})$ and $\kappa_4/n$ is the fourth standardized cumulant of $X$ upto the order of $O(1/n)$.

**Remark 1.2.** If the baseline hazard function $\lambda_0$ is unspecified, the constants $A, F$ and $\kappa_4$ in terms of $\sigma^2, \Delta, \zeta$ and $\mu_2 - \mu_9$ can be estimated easily from data. First we can estimate $\alpha_k, \alpha_k^{(i)}, \Lambda_k(t)$, and other quantities in (1.8) (see Gu (1992a)); then we can estimate $\sigma^2, \Delta, \zeta$ and $\mu_2 - \mu_9$. Finally, we can obtain consistent estimators for $A, F$ and $\kappa_4$.

**Remark 1.3.** If $\beta_0 = 0$, as in most applications, the expressions of $A, F$ and $\kappa_4$ in Theorem 1.1 can be simplified. In this case, we have $\alpha_k^{(i)}(t) = \alpha_k(t)$, then $\theta_k^{(i)}(t) = \theta_k(t)$. If further $z = 0$ or 1, great simplification can be achieved. In this case, $w$ can be considered as an alternative to the log-rank statistics (Mantel (1967)) and, compared with the normal approximation to the log-rank statistics, Theorem 1.1 provides more accurate approximations. Note that if we denote $p$ as the proportion of 1's among the $z$'s, we have $\eta(t) = -\theta_2(t)\Lambda_0(t)$ in (1.8) and $\alpha_k(t)/\alpha_0(t) = p$, for $k \geq 1$ and $\theta_2(t) = p(1-p), \eta_1(t) = 0, \theta_k(t) = p(1-p)^2$ for $k \geq 3$ and $\eta_k = p(1-p)\Lambda_0$ for $k \geq 2$. More specifically, we have

$$\begin{aligned} \sigma^2 &= p(1-p)\Sigma_0, \qquad \Delta = -p(1-p)(1-2p)\Sigma_0, \\ \zeta &= -p(1-p)(1-2p)\Sigma_1, \qquad \mu_2 = P(p)\Sigma_0, \\ \mu_3 &= P(p)\Sigma_1, \qquad \mu_4 = -Q(p)\Sigma_1, \\ \mu_6 &= P(p)\Sigma_2, \qquad \mu_8 = Q(p)\Sigma_2, \end{aligned}$$

and $\mu_5 = -\mu_4$, $\mu_7 = \mu_3$, $\mu_9 = -\mu_8/2$, where $P(p) = p(1-p)(1-3p+3p^2)$, $Q(p) = p^2(1-p)^2$ and for $k = 0, 1, 2$

$$\Sigma_k = \int_0^T \Lambda_0^k(t)\alpha_0(t)\lambda_0(t)dt.$$

**Example.** Consider the case $\beta_0 = 0$, $z = 1$ or 0 with $\Pr\{z = 1\} = p = 1/2$. First we have $\Delta = 0$, $\zeta = 0$. Further calculation shows that

$$A = 0, \quad F = \frac{3}{4} + \frac{-\Sigma_0 - 8\Sigma_1 + 8\Sigma_2}{4\Sigma_0^2}, \quad \kappa_4 = 0.$$

Numerical comparisons between the adjusted probability of (1.5), the probability using $\chi_1^2$ and the true probability using simulation (number of simulations $= 10{,}000{,}000$) with parameter $T = 2.5$ and $T = 4$, $\lambda_0(t) = 1$ are presented in Tables 1 and 2. No censoring is introduced. From the tables, we see that the relative errors of the adjusted probabilities are much better than those with the $\chi^2$ approximations, especially when the sample size is larger than 30.

Table 1. Limited comparisons between the adjusted, $\chi_1^2$ and the true probability $\Pr\{w \le u\}, T = 2.5$.

| | $u$ | 1 | 2 | 3 | 4 | 5 | 6.5 | 7.5 |
|---|---|---|---|---|---|---|---|---|
| $n = 10$ | true | .6488 | .8102 | .8891 | .9364 | .9601 | .9841 | .9920 |
| | adjusted | .6597 | .8230 | .9021 | .9442 | .9677 | .9855 | .9914 |
| $n = 30$ | true | .6722 | .8335 | .9096 | .9492 | .9710 | .9872 | .9925 |
| | adjusted | .6750 | .8361 | .9118 | .9511 | .9723 | .9880 | .9930 |
| $n = 50$ | true | .6772 | .8376 | .9129 | .9517 | .9726 | .9881 | .9931 |
| | adjusted | .6781 | .8388 | .9138 | .9524 | .9733 | .9885 | .9933 |
| | $\chi^2$ | .6827 | .8427 | .9167 | .9545 | .9747 | .9892 | .9938 |

Table 2. Limited comparisons between the adjusted, $\chi_1^2$ and the true probability $\Pr\{w \le u\}, T = 4$.

| | $u$ | 1 | 2 | 3 | 4 | 5 | 6.5 | 7.5 |
|---|---|---|---|---|---|---|---|---|
| $n = 10$ | true | .6431 | .8095 | .8889 | .9366 | .9605 | .9842 | .9921 |
| | adjusted | .6398 | .8059 | .8894 | .9354 | .9617 | .9822 | .9893 |
| $n = 30$ | true | .6685 | .8306 | .9079 | .9483 | .9706 | .9871 | .9924 |
| | adjusted | .6684 | .8304 | .9076 | .9481 | .9703 | .9869 | .9923 |
| $n = 50$ | true | .6739 | .8349 | .9109 | .9506 | .9720 | .9878 | .9929 |
| | adjusted | .6741 | .8353 | .9113 | .9507 | .9721 | .9878 | .9929 |
| | $\chi^2$ | .6827 | .8427 | .9167 | .9545 | .9747 | .9892 | .9938 |

## 2. Proofs

The proof is based on a series of representations. These representations will be stated as Lemmas below. The following functions are defined in Gu (1992a) and will be used in the proof.

$$g(y_i) = \int_0^T \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)}\right) dM_i(t),$$

$$h(y_i) = h_1(y_i) + h_2(y_i),$$

$$h_1(y_i) = \int_0^T \theta_2(t) dM_i(t), \quad h_2(y_i) = \int_0^T \rho_i(t)\lambda_0(t)dt,$$

$$\rho_i(t) = \xi_{i,2}(t) - \alpha_2(t) - 2\frac{\alpha_1(t)}{\alpha_0(t)}(\xi_{i,1}(t) - \alpha_1(t)) + \frac{\alpha_1^2(t)}{\alpha_0^2(t)}(\xi_{i,0}(t) - \alpha_0(t)), \quad (2.1)$$

$$\psi_0(y_i, y_j) = -\int_0^T \pi_i(t)dM_j(t), \quad \psi(y_i, y_j) = \psi_0(y_i, y_j) + \psi_0(y_j, y_i),$$

$$\pi_i(t) = \frac{1}{\alpha_0(t)}(\xi_{i,1}(t) - \alpha_1(t)) - \frac{\alpha_1(t)}{\alpha_0^2(t)}(\xi_{i,0}(t) - \alpha_0(t)).$$

$$\Psi(y_i, y_j) = \psi(y_i, y_j) - \frac{1}{\sigma^2}(g(y_i)h(y_j) + g(y_j)h(y_i)) + \frac{\Delta}{\sigma^4}g(y_i)g(y_j),$$

where $\sigma$ and $\Delta$ have been defined in (1.9).

In Lemma 2.1 $U_n = (\partial/\partial\beta)l_n(\beta)|_{\beta=\beta_0}$.

**Lemma 2.1.** *Under the conditions stated in the previous section, we have*

$$U_n(\beta_0) = \tilde{U}_n(\beta_0) + R_{1,n},$$

*where $R_{1,n}$ is such that for any $p \geq 2, E[|R_{1,n}|^p] = O(n^{-p})$ and*

$$\tilde{U}_n(\beta_0) = \sum_{i=1}^n g(y_i) + \frac{1}{n}\sum_{i<j}\psi(y_i, y_j) + \frac{1}{n}\sum_{i=1}^n \psi_0(y_i) + \frac{1}{n^2}\sum_{i<j<k} B(y_i, y_j, y_k), \quad (2.2)$$

*where $g(y_i)$, $\psi(y_i, y_j)$, and $\psi_0(y_i)$ are defined in* (2.1) *and*

$$\begin{aligned} B(y_i, y_j, y_k) &= \sum_{(i_1,i_2,i_3)} B_0(y_{i_1}, y_{i_2}, y_{i_3}), \\ B_0(y_i, y_j, y_k) &= -\int_0^T \Pi_{i,j}(t)dM_k(t), \\ \Pi_{i,j}(t) &= -\frac{\xi_{i,0}(t) - \alpha_0(t)}{\alpha_0(t)}\pi_j(t), \end{aligned} \quad (2.3)$$

*and where $\xi_{i,k}(t) = z_i^k e^{\beta z_i} 1_{[x_i \geq t]}$ and the summation in the first expression is over all premutations of $\{i, j, k\}$.*

**Proof of Lemma 2.1.** We start with the well known identity (See, for example, Gill (1984)):

$$U_n(\beta_0) = \sum_{i=0}^n \int_0^T \left(z_i - \frac{\hat{\alpha}_1(t)}{\hat{\alpha}_0(t)}\right) dM_i(t), \quad (2.4)$$

where $\hat{\alpha}_k(t) = \sum_{i=1}^n z_i^k \exp(\beta_0 z_i) 1_{[x_i \geq t]}$, $\alpha_k(t) = E[\hat{\alpha}_k(t)]$ and $M_i(t) = 1_{[x_i \leq t, \delta_i = 1]} - \int_0^t \exp(\beta_0 z_i) 1_{[x_i \geq s]} \lambda_0(t)dt$, $i = 1, \ldots, n$ are i.i.d. martingales. Using the identity

$$\begin{aligned} \frac{x}{y} - \frac{x_0}{y_0} &= \frac{x - x_0}{y_0} - \frac{x_0(y - y_0)}{y_0^2} + \frac{x_0(y - y_0)^2}{y_0^3} - \frac{(y - y_0)(x - x_0)}{y_0^2} \\ &\quad + O((x - x_0)^3 + (y - y_0)^3) \end{aligned}$$

we can write $U_n(\beta_0)-\tilde{U}_n(\beta_0)$ as $U$-statistics of degree four with degenerate kernel. Noting that the kernel of the $U$-statistics is bounded under our assumption, the lemma then follows from a lemma of Callaert and Janssen (1978).

The proofs of Lemmas 2.2–2.4 are similar to Lemma 2.1. We omit the proofs. In Lemma 2.2 we let $I_n = -(\partial^2/\partial\beta^2)l_n(\beta)|_{\beta=\beta_0}$.

**Lemma 2.2.** *Under the conditions stated in Section 1, we have*

$$I_n(\beta_0) = \tilde{I}_n(\beta_0) + R_{2,n}$$

*where $R_{2,n}$ is such that for any $p \geq 2$, $E[|R_{2,n}|^p] = O(n^{-p/2})$,*

$$\begin{aligned}
\tilde{I}_n(\beta_0) &= n\sigma^2 + \sum_{i=1}^{n} h(y_i) - \frac{1}{n}\sum_{i<j} C(y_i, y_j) + \gamma, \\
\gamma &= -\int_0^T \theta_2^{(1)}(t)\lambda_0(t)dt, \\
C(y_i, y_j) &= D(y_i, y_j) + D(y_j, y_i) - E(y_i, y_j) - E(y_j, y_i), \\
D(y_i, y_j) &= \int_0^T J_{i,j}(t)\lambda_0(t)dt, \\
J_{i,j}(t) &= \frac{\alpha_1^2(t)}{\alpha_0^3(t)}(\xi_{i,0}(t) - \alpha_0(t))(\xi_{j,0}(t) - \alpha_0(t)) \qquad (2.5) \\
&\quad - 2\frac{\alpha_1(t)}{\alpha_0^2(t)}(\xi_{i,0}(t) - \alpha_0(t))(\xi_{j,1}(t) - \alpha_1(t)) \\
&\quad + \frac{1}{\alpha_0(t)}(\xi_{i,1}(t) - \alpha_1(t))(\xi_{j,1}(t) - \alpha_1(t)), \\
E(y_i, y_j) &= \int_0^T \left[\frac{\rho_i(t)}{\alpha_0(t)} - \theta_2(t)\frac{(\xi_{i,0}(t) - \alpha_0(t))}{\alpha_0(t)}\right] dM_j(t),
\end{aligned}$$

*and where $h(y_i)$, $\rho_i(t)$ are defined in (2.1).*

Now, we represent $\Delta_n(\beta) = -(d/d\beta)I_n(\beta)$.

**Lemma 2.3.** *Under the conditions stated in Section 1, we have*

$$\Delta_n(\beta_0) = \tilde{\Delta}_n(\beta_0) + R_{3,n},$$

*where $R_{3,n}$ is such that for any $p \geq 2$, $E[|R_{3,n}|^p] = O(1)$ and*

$$\begin{aligned}
\tilde{\Delta}_n(\beta_0) &= n\Delta - \sum_{i=1}^{n} a(y_i) + O(1), \\
a(y_i) &= d(y_i) + b(y_i),
\end{aligned}$$

$$\begin{aligned} d(y_i) &= \int_0^T \left(\theta_3(t) - \frac{\alpha_1(t)}{\alpha_0(t)}\theta_2(t)\right) dM_i(t), \\ b(y_i) &= \int_0^T \left(\xi_{i,3}(t) - \alpha_3(t) - 3\frac{\alpha_1(t)}{\alpha_0(t)}(\xi_{i,2}(t) - \alpha_2(t)) \right. \\ &\quad - 3\left(\theta_2(t) - \frac{\alpha_1^2(t)}{\alpha_0^2(t)}\right)(\xi_{i,1}(t) - \alpha_1(t)) \\ &\quad \left. + \left(3\frac{\alpha_1(t)}{\alpha_0(t)}\theta_2(t) - \frac{\alpha_1^3(t)}{\alpha_0^3(t)}\right)(\xi_{i,0}(t) - \alpha_0(t))\right)\lambda_0(t)dt, \end{aligned} \tag{2.6}$$

*where* $\Delta$, $\theta_2$ *have been defined in* (1.9).

Finally, we have

**Lemma 2.4.** *Under the conditions stated in Section 1, we have*

$$\Delta_n'(\beta_0) = (d/d\beta)\Delta_n(\beta)|_{\beta=\beta_0} = n\Delta' + R_{4,n},$$

*where* $R_{4,n}$ *is such that for any* $p \geq 2$, $E[|R_{4,n}|^p] = O(n^{p/2})$ *and*

$$\Delta' = -\int_0^T \left(\alpha_4(t) - 4\frac{\alpha_3(t)\alpha_1(t)}{\alpha_0(t)} - 3\frac{\alpha_2^2(t)}{\alpha_0(t)} + 12\frac{\alpha_2(t)\alpha_1^2(t)}{\alpha_0^2(t)} - 6\frac{\alpha_1^4(t)}{\alpha_0^3(t)}\right)\lambda_0(t)dt. \tag{2.7}$$

**Proof of Theorem 1.1.** To simplify notation, we shall denote the terms that satisfy the property of $R_{1,n}$ in Lemma 2.1 as $O(n^{-1})$ and terms which satisfy the property of $R_{2,n}$ as $O(n^{-1/2})$. With this notation, we have $R_{3,n} = O(1)$ and $R_{4,n} = O(n^{1/2})$. From Lemma 4.4 of Gu (1992a), we have $\tilde{\beta} - \beta_0 = O(n^{-1/2})$, $\tilde{\beta} - \beta_0 = \Theta_1 + O(n^{-1})$, and $\hat{\beta} - \beta_0 = \Theta_2 + O(n^{-3/2})$, where

$$\begin{aligned} \Theta_1 &= \frac{1}{n\sigma^2}\sum_{i=1}^n g(y_i), \\ \Theta_2 &= \frac{1}{n\sigma^2}\sum_{i=1}^n g(y_i) + \frac{1}{n^2\sigma^2}\sum_{i<j}\Psi(y_i, y_j) + \frac{\Delta/2 - \zeta}{n\sigma^4}, \end{aligned} \tag{2.8}$$

and where $\Psi$ was defined in (2.1), $\Delta$, $\zeta$ and $\sigma$ were defined in (1.9). By (1.2), Lemma 2.1–Lemma 2.4 and three terms Taylor expansion we can show that

$$w = \tilde{I}_n(\beta_0)(\hat{\beta} - \beta_0)^2 - \frac{2}{3}\tilde{\Delta}_n(\beta_0)(\hat{\beta} - \beta_0)^3 - \frac{1}{4}n\Delta'(\hat{\beta} - \beta_0)^4 + O(n^{-3/2}). \tag{2.9}$$

To handle the first term on the right, let us define

$$\Theta = \tilde{U}_n(\beta_0) + \frac{1}{2}\tilde{\Delta}_n(\beta_0)\Theta_2^2 + \frac{1}{6}n\Delta'\Theta_1^3.$$

By (2.8), Lemma 2.1–Lemma 2.4 and three terms Taylor expansion we can show that $\hat{\beta} - \beta_0 = \tilde{I}_n^{-1}(\beta_0)\Theta + O(n^{-2})$. Therefore the first term on the right of (2.9) can be written as $\tilde{I}_n^{-1}(\beta_0)\Theta^2 + O(n^{-3/2})$. In the subsequent calculation, $\tilde{I}_n^{-1}(\beta_0)$ is replaced by $I_{0,n}^{-1}$, where

$$I_{0,n}^{-1} = \frac{1}{n\sigma^2} - \frac{1}{n^2\sigma^4}\sum_{i=1}^{n} h(y_i) + \frac{1}{n^3\sigma^4}\sum_{i<j} C(y_i, y_j) - \frac{1}{n^2\sigma^4}\gamma.$$

It is easy to see that $\tilde{I}_n^{-1}(\beta_0) = I_{0,n}^{-1} + O(n^{-5/2})$. For the second and third terms on the right of (2.9), using $\Theta_1$ and $\Theta_2$ of (2.8) was enough. After some further algebraic manipulation, we have

$$w = \left(M + \frac{N}{\sqrt{n}}\right)^2 + \frac{2QM}{n} + O(n^{-3/2}), \tag{2.10}$$

where

$$\begin{aligned}
M &= \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} g(y_i),\\
N &= \frac{\Delta}{6\sigma^3}M^2 + \frac{1}{n\sigma}\sum_{i<j}\psi(y_i, y_j) - \frac{1}{2\sigma^2}M\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(y_i)\right),\\
Q &= \left(\frac{\Delta'}{24\sigma^4} + \frac{\Delta^2}{9\sigma^6}\right)M^3 + \frac{1}{\sigma}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_0(y_i)\right) - \frac{\gamma}{2\sigma^2}M \\
&\quad + \frac{1}{\sigma}\left(\frac{1}{n^{3/2}}\sum_{i<j<k} B(y_i, y_j, y_k)\right) + \frac{\Delta}{3\sigma^4}M\left(\frac{1}{n}\sum_{i<j}\psi(y_i, y_j)\right) \\
&\quad - M^2\left(\frac{1}{6\sigma^3}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} a(y_i) + \frac{5\Delta}{12\sigma^5}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(y_i)\right) \\
&\quad + \frac{1}{2\sigma^2}M\left(\frac{1}{n}\sum_{i<j} C(y_i, y_j)\right) - \frac{1}{2\sigma^3}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(y_i)\right)\left(\frac{1}{n}\sum_{i<j}\psi(y_i, y_j)\right).
\end{aligned} \tag{2.11}$$

From (2.10), $N = O(1)$ and $Q = O(1)$, so that we have

$$w = X^2 + O(n^{-3/2}), \tag{2.12}$$

where

$$X = M + \frac{N}{\sqrt{n}} + \frac{Q}{n}. \tag{2.13}$$

It is easy to see that

$$E[M] = 0, \quad E[N] = A, \quad \text{and} \quad E[Q] = 0 + O(n^{-1/2}), \tag{2.14}$$

where $A$ has been defined in (1.6) and therefore $E[X] = A/\sqrt{n} + O(n^{-3/2})$.

The next step is to write $X - E[X]$ as an $U$-statistics of degree three. We have

$$X - E[X] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G(y_i) + \frac{1}{n^{3/2}}\sum_{i<j} S(y_i, y_j) + \frac{1}{n^{5/2}}\sum_{i<j<k} T(y_i, y_j, y_k) + O(n^{-3/2}), \tag{2.15}$$

where

$$\begin{aligned} G(y_i) &= \frac{g(y_i)}{\sigma} + \frac{1}{n}\left(\frac{\Delta'}{8\sigma^4} + \frac{1\Delta^2}{3\sigma^6} - \frac{E[g(y_1)a(y_1)]}{3\sigma^4} - \frac{5\Delta\zeta}{6\sigma^6} - \frac{\gamma}{2\sigma^2}\right)\frac{g(y_i)}{\sigma} \\ &\quad + \frac{\Delta}{6n\sigma^5}(g^2(y_i) - \sigma^2) - \frac{1}{2n\sigma^3}(g(y_i)h(y_i) - \zeta) + \frac{1}{n\sigma}\psi_0(y_i) \\ &\quad + \frac{1}{2n\sigma^3}(H(y_i) - K(y_i)) - \frac{a(y_i)}{6n\sigma^3} - \frac{5\Delta}{12n\sigma^5}h(y_i) \\ S(y_i, y_j). &= \frac{\Delta}{3\sigma^5}g(y_i)g(y_j) + \frac{1}{\sigma}\psi(y_i, y_j) - \frac{1}{2\sigma^3}[h(y_i)g(y_j) + h(y_j)g(y_i)], \\ T(y_i, y_j, y_k) &= \left(\frac{\Delta'}{4\sigma^7} + \frac{2\Delta^2}{3\sigma^9}\right)g(y_i)g(y_j)g(y_k) + \frac{1}{\sigma}B(y_i, y_j, y_k) \qquad (2.16) \\ &\quad + \sum_{\{i_1\},\{j_1,j_2\}}\left[\frac{\Delta}{3\sigma^5}g(y_{i_1})\psi(y_{j_1}, y_{j_2}) - \frac{1}{3\sigma^5}a(y_{i_1})g(y_{j_1})g(y_{j_2})\right. \\ &\quad + \frac{1}{2\sigma^3}g(y_{i_1})C(y_{j_1}, y_{j_2}) - \frac{1}{2\sigma^3}h(y_{i_1})\psi(y_{j_1}, y_{j_2}) \\ &\quad \left. - \frac{5\Delta}{6\sigma^7}h(y_{i_1})g(y_{j_1})g(y_{j_2})\right], \end{aligned}$$

where the summation in the last expression is over all three possible ways of grouping $i$, $j$, $k$ into two groups $\{i_1\}$ and $\{j_1, j_2\}$ and

$$\begin{aligned} H(y_i) &= E[g(y_j)C(y_i, y_j)|y_i] \\ K(y_i) &= E[h(y_j)\psi(y_i, y_j)|y_i]. \end{aligned} \tag{2.17}$$

Next we shall use the Edgeworth expansion result for $U$-statistics developed by Gu (1992c). In order to use the Edgeworth expansion result for $U$-statistics, we need to calculate the variance of $X$ up to $O(n^{-1})$, $\kappa_3$ up to $O(n^{-1/2})$ and $\kappa_4$ up to $O(1)$. From (2.15), $\mathrm{Var}(X) = E[G^2(y_1)] + E[S^2(y_1, y_2)]/(2n) + O(n^{-3/2})$. Ignoring the parts of order $n^{-3/2}$ and higher, after some lengthy calculations, we get

$$\mathrm{Var}(X) = 1 + F/n,$$

where

$$F = -\frac{11\Delta\zeta}{6\sigma^6} + \frac{7\Delta^2}{18\sigma^6} + \frac{\zeta^2}{4\sigma^6} + \frac{\Delta'}{4\sigma^4} - \frac{E[g(y_1)a(y_1)]}{\sigma^4} - \frac{E[g^2(y_1)h(y_1)]}{\sigma^4}$$

$$+\frac{E[h^2(y_1)]}{4\sigma^4}+\frac{E[g(y_1)g(y_2)C(y_1,y_2)]}{\sigma^4}-\frac{2E[\psi(y_1,y_2)h(y_1)g(y_2)]}{\sigma^4}$$
$$+\frac{E[\psi^2(y_1,y_2)]}{2\sigma^2}+\frac{2E[g(y_1)\psi_0(y_1)]}{\sigma^2}-\frac{\gamma}{\sigma^2}. \quad (2.18)$$

Simple algebra using (2.16) shows that

$$\kappa_3=(\mathrm{Var}(X))^{-3/2}\Big\{E[G^3(y_1)]+3E[G(y_1)G(y_2)S(y_1,y_2)]\Big\}=O(n^{-1}),$$

and therefore $\kappa_3$ can be considered as zero. The next task is to calculate $\kappa_4$ using (2.16). We have

$$\begin{aligned}\kappa_4 &= \sigma_G^{-4}\{E[G^4(y_1)]-3\sigma_G^4+12E[G^2(y_1)G(y_2)S(y_1,y_2)]\\ &\quad +4E[T(y_1,y_2,y_3)G(y_1)G(y_2)G(y_3)]+12E[S(y_1,y_2)S(y_1,y_3)G(y_2)G(y_3)]\}\\ &= \sigma^{-4}\Big\{E[g^4(y_4)]-3\sigma^4-\frac{9\zeta^2}{\sigma^2}+\Delta'\\ &\quad +12E[g^2(y_1)g(y_2)\psi(y_1,y_2)]-6E[g^2(y_1)h(y_1)]-4E[a(y_1)g(y_1)]\\ &\quad +6E[C(y_1,y_2)g(y_1)g(y_2)]+12E[\psi(y_1,y_2)\psi(y_1,y_3)g(y_2)g(y_3)]\\ &\quad -12E[h(y_1)g(y_2)\psi(y_1,y_2)]+3E[h^2(y_1)]\Big\}, \end{aligned} \quad (2.19)$$

where we have used $E[g^3(y_1)]=3\zeta-\Delta$, $E[g(y_1)h(y_1)]=\zeta$, $E[g(y_1)g(y_2)\psi(y_1,y_2)]=0$, and $E[B(y_1,y_2,y_3)g(y_1)g(y_2)g(y_3)]=0$ (see Gu (1992b)).

Using the result on the Edgeworth expansion of $U$-statistics of degree three of Gu (1992c), the distribution function of $(X-E[X])/\sqrt{\mathrm{Var}(X)}$ is

$$F_n(x)=\Phi(x)-\frac{\kappa_4}{24n}\phi(x)(x^3-3x)+O(n^{-3/2}). \quad (2.20)$$

Since $E[X]=A/\sqrt{n}$ (2.14), after a change of variable, we get the distribution function of $X^2$ (and therefore $w$) as the right side of (1.5). The proof is finished if we can show the right side of (2.18) equals that of (1.6) and the right side of (2.19) equals that of (1.7). To demonstrate this, let us consider, for example, $E[g(y_1)\psi_0(y_1)]$. From the definition, we have

$$\begin{aligned}E[g(y_1)\psi_0(y_1)] &= -E\Big[\int_0^T\Big(z_1-\frac{\alpha_1(t)}{\alpha_0(t)}\Big)dM_1(t)\int_0^T\pi_1(t)dM_1(t)\Big]\\ &= -E\Big[\int_0^T\Big(z_1-\frac{\alpha_1(t)}{\alpha_0(t)}\Big)\pi_1(t)\xi_{1,0}(t)\lambda_0(t)dt\Big]\\ &= -\int_0^T E\Big[\Big(\xi_{1,1}(t)-\frac{\alpha_1(t)}{\alpha_0(t)}\xi_{1,0}(t)\Big)^2\Big]\frac{\lambda_0(t)}{\alpha_0(t)}dt=\gamma, \end{aligned} \quad (2.21)$$

since

$$E\left[\left(\xi_{1,1}(t)-\frac{\alpha_1(t)}{\alpha_0(t)}\xi_{1,0}(t)\right)^2\right]=\theta_2^{(1)}(t).$$

Similar calculations show that

$$\begin{aligned}
E[\psi^2(y_1,y_2)] &= -2\gamma,\\
E[h^2(y_1)] &= -2\mu_5+2\mu_7+\nu-\sigma^4,\\
E[g(y_1)a(y_1)] &= -\mu_3-3\mu_4,\\
E[g^2(y_1)h(y_1)] &= 2\mu_4-\mu_5+\mu_6+\mu_7+\nu-\sigma^4,\\
E[g(y_1)g(y_2)C(y_1,y_2)] &= 2\mu_8, \qquad (2.22)\\
E[\psi(y_1,y_2)h(y_1)g(y_2)] &= -\mu_4+\mu_9,\\
E[g^4(y_4)] &= \mu_2-4\mu_3+6\mu_6,\\
E[\psi(y_1,y_2)\psi(y_1,y_3)g(y_2)g(y_3)] &= \mu_8,\\
E[g^2(y_1)g(y_2)\psi(y_1,y_2)] &= -\mu_4-2\mu_8+\mu_9,
\end{aligned}$$

where $\mu_2$-$\mu_9$ have been defined in (1.9), $\gamma$ has been defined in (2.5) and

$$\nu=\int_0^T\theta_2^2(t)\alpha_0(t)\lambda_0(t)dt.$$

Substituting the quantities on the right of (2.21) and (2.22) into (2.18) and (2.19), also noting that $\Delta'=-\mu_2+3\mu_9$, we get (1.6) and (1.7).

## Acknowledgement

Minggao Gu is partially supported by Natural Science and Engineering research Council of Canada and Fonds pour la Formation de Chercheurs et L'Aide à la Recherche. The authors are grateful for the remarks made by the editors and referees, which led to an improvement of an earlier version of this paper.

## References

Andersen, P. K. and Gill, D. R. (1982). Cox's regression model for counting processes. *Ann. Statist.* **10**, 1100–1120.

Bailey, K. R. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the Cox regression model. *Ann. Statist.* **11**, 39–48.

Barndorff-Nielsen, O. E. and Cox, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser.B* **46**, 483–495.

Bartlett, M. S. (1937). Properties of sufficiency and statistical test. *Proc. Roy. Soc. Ser.A* **160**, 268–282.

Cox, D. R. (1972). Regression model and life-tables (with discussion). *J. Roy. Statist. Soc. Ser.B* **34**, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for $U$-statistics. *Ann. Statist.* **6**, 417–421.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics* **23**, 65–78.

Gill, R. D. (1984). Understanding Cox's regression model: A martingale approach. *J. Amer. Statist. Assoc.* **79**, 441–447.

Gu, M. (1992a). On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20**, 399–414.

Gu, M. (1992b). Some moments concerning the Cox regression model. Report from the Dept. of Math. and Statist., McGill University 92–01.

Gu, M. (1992c). On the Edgeworth expansion for $U$-statistics. Report from the Dept. of Math. and Statist., McGill University 92–03.

Naes, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Statist.* **9**, 107–115.

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

Department of Mathematics and Statistics, McGill University, Montreal, Quebec H3A 2K6, Canada.

Department of Statistics and Operation Research, Fudan University, Shanghai, 200433.

(Received November 1991; accepted February 1993)