The Robust Inference for the Cox Proportional Hazards Model

D. Y. LIN and L. J. WEI*

We derive the asymptotic distribution of the maximum partial likelihood estimator $\hat{\beta}$ for the vector of regression coefficients β under a possibly misspecified Cox proportional hazards model. As in the parametric setting, this estimator $\hat{\beta}$ converges to a well-defined constant vector β^*. In addition, the random vector $n^{1/2}(\hat{\beta} - \beta^*)$ is asymptotically normal with mean 0 and with a covariance matrix that can be consistently estimated. The newly proposed robust covariance matrix estimator is similar to the so-called "sandwich" variance estimators that have been extensively studied for parametric cases. For many misspecified Cox models, the asymptotic limit β^* or part of it can be interpreted meaningfully. In those circumstances, valid statistical inferences about the corresponding covariate effects can be drawn based on the aforementioned asymptotic theory of $\hat{\beta}$ and the related results for the score statistics. Extensive studies demonstrate that the proposed robust tests and interval estimation procedures are appropriate for practical use. In particular, the robust score tests perform quite well even for small samples. In contrast, the conventional model-based inference procedures often lead to tests with supranominal size and confidence intervals with rather poor coverage probability.

KEY WORDS: Asymptotic theory; Model misspecification; Partial likelihood; Regression; Robustness; Survival data.

1. INTRODUCTION

The Cox proportional hazards model assumes that the hazard function $\lambda(t)$ for the failure time T of an individual with a p vector of covariates $Z(t) = (Z_1(t), \ldots, Z_p(t))'$ has the following form:

$$\lambda(t; Z) = \lambda_0(t)\exp\{\beta' Z(t)\}, \qquad (1.1)$$

where $\beta = (\beta_1, \ldots, \beta_p)$ is a p vector of unknown regression coefficients and $\lambda_0(t)$ is an unspecified baseline hazard function. Let $X_1, \ldots, X_n$ be n possibly right-censored failure times, and let $Z_1, \ldots, Z_n$ be the corresponding covariate vectors, where Z_i is observed on $[0, X_i]$. The censoring is assumed to be noninformative. Then the maximum partial likelihood estimator $\hat{\beta}$ is the value that maximizes the partial likelihood function (Cox 1972, 1975)

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\{\beta' Z_i(X_i)\}}{\sum_{j\in\mathcal{R}_i} \exp\{\beta' Z_j(X_i)\}}\right]^{\delta_i}, \qquad (1.2)$$

where $\mathcal{R}_i$ is the set of labels attached to the individuals at risk at time X_i-, and $\delta_i = 1$ if X_i is an observed failure time and $\delta_i = 0$ otherwise.

Now, let $\hat{A}(\beta) = -n^{-1}\partial^2 \log L(\beta)/\partial\beta^2$. If the assumed Model (1.1) is correct, then $n^{1/2}(\hat{\beta} - \beta)$ converges in distribution to a p-dimensional normal vector with mean 0 and with a covariance matrix that can be consistently estimated by $\hat{A}^{-1}(\hat{\beta})$ (see Andersen and Gill 1982).

The consequences of misspecifying the Cox model have been extensively investigated in recent years (see Gail, Wieand, and Piantadosi 1984; Lagakos 1988; Lagakos and Schoenfeld 1984; Morgan 1986; O'Neill 1986; Solomon 1984; Struthers and Kalbfleisch 1986). When the assumed Model (1.1) is incorrect, the estimator $\hat{\beta}$ converges to a well-defined constant vector β^* (see Struthers and Kalbfleisch 1986). In many cases, β^* or part of it can be interpreted meaningfully. For example, suppose that the true hazard function takes the form $\lambda_1(t)\exp(\gamma Z^2)$, where Z is symmetric about 0, γ is a constant, and $\lambda_1(t)$ is a baseline hazard function. Let the assumed model be $\lambda_0(t)\exp(\beta Z)$. Then it can be shown that the asymptotic limit β^* of $\hat{\beta}$ is 0 provided that the censoring of responses acts independently of Z. In this case, it is possible to construct a valid test based on $\hat{\beta}$ for testing the null hypothesis that there is no linear effect of Z on the failure time. As will be shown in Section 3, however, $\hat{A}^{-1}(\hat{\beta})$ does not provide a proper variance estimator of $n^{1/2}\hat{\beta}$.

In the *parametric* setting, a number of techniques have been suggested for handling misspecified models (e.g., Gail, Tan, and Piantadosi 1988; Huber 1967; Kent 1982; Royall 1986; White 1982). As an illustration, suppose that we are interested in testing the null hypothesis H_0 of no treatment effect in a randomized clinical trial comparing a treatment and a control. Let τ and ψ denote, respectively, the parameters representing the effects of treatment and covariates in a working parametric model. In addition, let $\theta = (\tau, \psi)$, and let $\hat{\theta}$ be the maximum likelihood estimator of θ from the log-likelihood function $l(\theta)$ under the working model. In addition, denote $\hat{A}(\theta) = -n^{-1}\partial^2 l(\theta)/\partial\theta^2$ and $\hat{B}(\theta) = n^{-1} \sum U_i(\theta)U_i'(\theta)$, where $U_i(\theta)$ is the contribution from the ith observation to the score function $U(\theta) = \partial l(\theta)/\partial\theta$. Then, under some mild regularity conditions, $\hat{\theta}$ converges to a well-defined constant vector $\theta^* = (\tau^*, \psi^*)$, and $n^{1/2}(\hat{\theta} - \theta^*)$ is asymptotically normal with zero mean and with a covariance matrix that can be consistently estimated by $\hat{V}(\hat{\theta}) = \hat{A}^{-1}(\hat{\theta})\hat{B}(\hat{\theta})\hat{A}^{-1}(\hat{\theta})$, the so-called "sandwich" estimator. For many misspecified parametric models, $\tau^* = 0$ under H_0. This is true, for example, when some relevant covariates are omitted from a generalized linear model (see Gail

* D. Y. Lin is Postdoctoral Research Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. L. J. Wei is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706. This work was supported by a Population Council Fellowship (for Lin) and National Institutes of Health Grant RO1-CA45544 (for Wei). The authors are grateful to the editor, an associate editor, and two referees for their helpful comments.

Journal of the American Statistical Association
December 1989, Vol. 84, No. 408, Theory and Methods

et al. 1984). In such situations, H_0 can be expressed as $\tau = 0$ in the working model.

Next, let $\hat{V}(\tau, \psi)$, $\hat{A}(\tau, \psi)$, $U_i(\tau, \psi)$, and $U(\tau, \psi)$ be partitioned according to the partition (τ, ψ) of θ; that is,

$$\hat{V}(\tau, \psi) = \begin{bmatrix} \hat{V}_{\tau\tau}(\tau, \psi) & \hat{V}_{\tau\psi}(\tau, \psi) \\ \hat{V}_{\psi\tau}(\tau, \psi) & \hat{V}_{\psi\psi}(\tau, \psi) \end{bmatrix}, \quad U_i(\tau, \psi) = \begin{bmatrix} U_{i\tau}(\tau, \psi) \\ U_{i\psi}(\tau, \psi) \end{bmatrix},$$

and so on. Then the robust Wald test and score test (also called C_α test) for testing H_0 are, respectively, $n\hat{\tau}^2/\hat{V}_{\tau\tau}(\hat{\tau}, \hat{\psi})$ and $U_\tau^2(0, \hat{\psi}_0)/\Sigma\ \{U_{i\tau}(0, \hat{\psi}_0) - \hat{A}_{\tau\psi}(0, \hat{\psi}_0)\hat{A}_{\psi\psi}^{-1}(0, \hat{\psi}_0)U_{i\psi}(0, \hat{\psi}_0)\}^2$, where $\hat{\psi}_0$ is the restricted maximum likelihood estimator of ψ given $\tau = 0$. It is interesting to note that the robust score test can be interpreted as a permutation test based on the residuals computed from the regression on covariates, but not on treatment (see Gail et al. 1988).

In this article, various robust procedures similar to those for parametric models are proposed and evaluated for Cox's *semiparametric* model. In Section 2, we derive the asymptotic distribution of the maximum partial likelihood estimator when Model (1.1) may be misspecified. The random vector $n^{1/2}(\hat{\beta} - \beta^*)$ is shown to be asymptotically normal with mean 0 and with a covariance matrix that generally differs from the asymptotic limit of $\hat{A}^{-1}(\hat{\beta})$. A consistent variance–covariance estimator is suggested. In Section 3, valid Wald and score statistics are presented for misspecified Cox models. Extensive empirical studies demonstrate the superior performance of the robust procedures over their model-based counterparts.

2. ASYMPTOTIC PROPERTIES OF THE MAXIMUM PARTIAL LIKELIHOOD ESTIMATOR

For the ith individual, let $\lambda_i(t)$ be the true hazard function, and $Y_i(t) = I\{X_i \geq t\}$, where $I\{\cdot\}$ is the indicator function. We assume throughout the article that (X_i, δ_i, Z_i) $(i = 1, \ldots, n)$ are n iid realizations of (X, δ, Z), that Z is bounded, and that the support of the failure time T properly contains that of the censoring variable.

It is convenient to introduce the following notation

$$S^{(r)}(t) = n^{-1} \sum_{i=1}^{n} Y_i(t)\lambda_i(t)Z_i(t)^{\otimes r}, \qquad s^{(r)}(t) = E\{S^{(r)}(t)\},$$

$$S^{(r)}(\beta, t) = n^{-1} \sum_{i=1}^{n} Y_i(t)\exp\{\beta' Z_i(t)\}Z_i(t)^{\otimes r},$$

and

$$s^{(r)}(\beta, t) = E\{S^{(r)}(\beta, t)\},$$

for $r = 0, 1, 2$, where for a column vector a, $a^{\otimes 2}$ refers to the matrix aa', $a^{\otimes 1}$ refers to the vector a, and $a^{\otimes 0}$ refers to the scalar 1 and the expectations are taken with respect to the true model of (X_i, δ_i, Z_i) $(i = 1, \ldots, n)$.

Let (1.1) be the working model for (X_i, δ_i, Z_i) $(i = 1, \ldots, n)$. The logarithm of the partial likelihood function (1.2) can be expressed as

$$l(\beta) = \sum_{i=1}^{n} \delta_i[\beta' Z_i(X_i) - \log\{S^{(0)}(\beta, X_i)\}].$$

The corresponding score function is

$$U(\beta) = \sum_{i=1}^{n} \delta_i \left\{ Z_i(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\}.$$

The maximum partial likelihood estimator $\hat{\beta}$ is the solution to the system of equations $U(\beta) = 0$.

The asymptotic properties of the maximum partial likelihood estimator $\hat{\beta}$ when the assumed Cox model (1.1) is valid have been studied, for example, by Tsiatis (1981), Andersen and Gill (1982), Næs (1982), and Bailey (1983). Here, we derive the asymptotic distribution of $n^{1/2}\hat{\beta}$ when Model (1.1) may be incorrect. By the techniques used in the proofs of lemma 3.1 and theorem 4.2 of Andersen and Gill (1982), one can easily show that $\hat{\beta}$ converges in probability to a p vector of constants β^*. Here, β^* is the unique solution to the system of equations

$$\int_0^\infty s^{(1)}(t)\, dt - \int_0^\infty \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} s^{(0)}(t)\, dt = 0, \qquad (2.1)$$

provided that $A(\beta^*)$ is positive definite, where

$$A(\beta) = \int_0^\infty \left\{ \frac{s^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - \frac{s^{(1)}(\beta, t)^{\otimes 2}}{s^{(0)}(\beta, t)^2} \right\} s^{(0)}(t)\, dt.$$

The foregoing result is a multivariate generalization of theorem 2.1 of Struthers and Kalbfleisch (1986). The positive-definiteness of $A(\beta^*)$ is assumed in the sequel. If Model (1.1) is correct, then $\beta^* = \beta$ and $\hat{A}(\hat{\beta})$ is a consistent estimator of $A(\beta)$.

Now, let

$$W_i(\beta) = \delta_i \left\{ Z_i(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right\} - \sum_{j=1}^{n} \frac{\delta_j Y_i(X_j)\exp\{\beta' Z_i(X_j)\}}{nS^{(0)}(\beta, X_j)} \times \left\{ Z_i(X_j) - \frac{S^{(1)}(\beta, X_j)}{S^{(0)}(\beta, X_j)} \right\}.$$

In addition, let $\hat{B}(\beta) = n^{-1} \Sigma\ W_i(\beta)^{\otimes 2}$, and let $\hat{V}(\beta) = \hat{A}^{-1}(\beta)\hat{B}(\beta)\hat{A}^{-1}(\beta)$. The large-sample properties of the maximum partial likelihood estimator $\hat{\beta}$ under a possibly misspecified Cox model are given in the following theorem.

Theorem 2.1. The random vector $n^{1/2}(\hat{\beta} - \beta^*)$ is asymptotically normal with mean 0 and with a covariance matrix that can be consistently estimated by $\hat{V}(\hat{\beta})$.

The proof of Theorem 2.1 is provided in the Appendix. Notice that the robust variance–covariance estimator $\hat{V}(\hat{\beta})$ for the Cox model is rather similar to its parametric counterpart $\hat{V}(\hat{\theta})$ described in Section 1. The key difference is that the matrix $\hat{B}(\beta)$ in $\hat{V}(\beta)$ takes a more complicated form than the corresponding $\hat{B}(\theta)$ in $\hat{V}(\theta)$ because the score function is not a sum of n iid random vectors in a partial likelihood setting.

3. ROBUST STATISTICAL INFERENCE

In this section, we demonstrate by several examples that the statistical procedures based on the maximum partial likelihood estimator $\hat{\beta}$ and the robust variance–covariance estimator $\hat{V}(\hat{\beta})$ are appropriate for practical use under a wide range of misspecified Cox models. On the other hand, the procedures with the model-based variance–covariance estimator $\hat{A}^{-1}(\hat{\beta})$ have some undesirable properties. The related score statistics are also studied.

Suppose that one is interested in testing the null hypothesis H_0 that the failure time T does not depend on $Z_1(t)$, the first component of the covariate vector $Z(t)$. When the assumed model (1.1) is true, a valid test statistic for H_0 can be constructed based on $\hat{\beta}_1$, with the variance estimator derived from either $\hat{A}^{-1}(\hat{\beta})$ or $\hat{V}(\hat{\beta})$. When Model (1.1) is false, a valid test for H_0 based on $\hat{\beta}_1$ is possible if the limit β_1^* of $\hat{\beta}_1$ is still 0 under H_0. Here, we provide two important examples of model misspecification in which this condition is satisfied. For simplicity, let us assume that there are no censored observations in the data. Recall that β^* is the unique solution to the system of p equations (2.1). Now, for $k = 1, \ldots, p$, let $s_k^{(1)}(\beta, t)$ and $s_k^{(1)}(t)$ be the kth components of $s^{(1)}(\beta, t)$ and $s^{(1)}(t)$, respectively. Then for $\beta_1^* = 0$, there exist $(p - 1)$ constants $\beta_2^*, \ldots, \beta_p^*$ that uniquely solve the following $(p - 1)$ equations under H_0:

$$\int_0^\infty s_k^{(1)}(t)\,dt - \int_0^\infty \frac{s_k^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} s^{(0)}(t)\,dt = 0, \qquad k = 2, \ldots, p.$$

Thus, to show that $(0, \beta_2^*, \ldots, \beta_p^*)$ is the solution to the system of p equations (2.1) under H_0, it is sufficient to verify that $\beta_1^* = 0$ satisfies the following condition under H_0:

$$\int_0^\infty s_1^{(1)}(t)\,dt - \int_0^\infty \frac{s_1^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} s^{(0)}(t)\,dt = 0. \tag{3.1}$$

In our first example, suppose that $Z_1(t)$ is independent of all other relevant covariates, some of which may be mistakenly omitted from the model. The true model need not take an exponential regression form, and neither does it have to be a proportional hazards model. Then, under H_0, $s_1^{(1)}(t) = s^{(0)}(t)E\{Z_1(t)\}$ and $s_1^{(1)}(\beta^*, t) = s^{(0)}(\beta^*, t)E\{Z_1(t)\}$ when $\beta_1^* = 0$; therefore, (3.1) is satisfied. In the second example, suppose that $Z_1(t)$ is symmetric about 0 and that $Z_1^2(t)$ has an important effect on the hazard function $\lambda(t)$ but is mistakenly omitted. Other relevant covariates are assumed to be independent of $Z_1(t)$ and may also be omitted from the model. The true model may have a nonexponential regression form, or it may be a nonproportional hazards model. Here, it is easy to show that for $\beta_1^* = 0$, $s_1^{(1)}(\beta^*, t) = s_1^{(1)}(t) = 0$ under H_0. Again, (3.1) is satisfied.

For these two examples of model misspecification, the Wald test for testing H_0 based on $\hat{\beta}_1$ and $\hat{V}(\hat{\beta})$ is asymptotically valid, whereas its model-based counterpart with $\hat{A}^{-1}(\hat{\beta})$ may not be. Now, for the corresponding score tests, let $U(\beta_1, \eta)$, $W_i(\beta_1, \eta)$, and $\hat{A}(\beta_1, \eta)$ be partitioned according to the partition (β_1, η) of β. In addition, let $\hat{\eta}_0$ denote the restricted maximum partial likelihood estimator of η given $\beta_1 = 0$. Then, the model-based and the robust score tests for H_0 are, respectively, $n^{-1}U_{\beta_1}^2(0, \hat{\eta}_0)/\{\hat{A}_{\beta_1\beta_1}(0, \hat{\eta}_0) - \hat{A}_{\beta_1\eta}(0, \hat{\eta}_0)\hat{A}_{\eta\eta}^{-1}(0, \hat{\eta}_0)\hat{A}'_{\beta_1\eta}(0, \hat{\eta}_0)\}$ and $U_{\beta_1}^2(0, \hat{\eta}_0)/\sum \{W_{i\beta_1}(0, \hat{\eta}_0) - \hat{A}_{\beta_1\eta}(0, \hat{\eta}_0)\hat{A}_{\eta\eta}^{-1}(0, \hat{\eta}_0)W_{i\eta}(0, \hat{\eta}_0)\}^2$. These procedures were derived from the arguments given in the Appendix and in Cox and Hinkley (1974, pp. 321–324).

Extensive empirical studies have been conducted to evaluate the properties of the model-based and the robust tests for practical sample sizes. Some results from these studies are displayed in Table 1. Rows 1–4 of the table are on the omission of relevant covariates from Cox models, rows 5–8 are on the misspecification of regression forms with possible omission of relevant covariates, and rows 9–12 are on nonproportional hazards models also with possible omission of relevant covariates. It is interesting to observe that the model-based Wald and score tests are comparable. These conventional tests seem to retain near nominal size when an independent covariate is omitted from a Cox model, which confirms the findings of Lagakos and Schoenfeld (1984). Their size, however, often well exceeds the nominal level when an uncorrelated but dependent covariate is omitted from a Cox model or when the true model is not in a proportional hazards form. One may notice that these tests can be rather conservative when the regression form is incorrect. In all of the cases studied here, the robust Wald and score tests maintain their size near the nominal level, especially for large samples. The robust score test tends to perform better than its Wald counterpart. Additional Monte Carlo studies not shown here have indicated that the behavior of the robust score test is fairly satisfactory, even for small samples such as $n = 20$. These findings suggest that the robust score test be used in practice.

Under certain misspecified Cox models, it is meaningful to make the quantitative inferences about the covariate effects on the failure time. For example, suppose that we work under Model (1.1) but the true model is of the form $\lambda(t) = \lambda_1(t)\exp(\gamma'Z + \xi'C)$, where the omitted covariate vector C is uncorrelated with Z. Then it is interesting to estimate the effect of, say Z_1, on the failure time. Obviously, an asymptotically valid confidence interval for the regression coefficient γ_1 of Z_1 can be constructed based on $\hat{\beta}_1$ and $\hat{V}(\hat{\beta})$ as long as $\beta_1^* = \gamma_1$. This condition is satisfied, for instance, in rows 1–4 of Table 1. Notice that the size of a Wald test is the complement of the coverage probability of the corresponding Wald-type confidence interval. Thus, in rows 2–4 of Table 1, the confidence intervals with $\hat{V}(\hat{\beta})$ have proper coverage probability, whereas those with $\hat{A}^{-1}(\hat{\beta})$ do not. It should be pointed out that β_1^* is generally unequal to γ_1 when $\gamma_1 \neq 0$. However, our extensive numerical studies not shown here have indicated that the difference between β_1^* and γ_1 is small unless $\|\xi\|$ is relatively large, which confirms the findings of Struthers and Kalbfleisch (1986). In addition, the confidence intervals based on $\hat{\beta}_1$ and $\hat{V}(\hat{\beta})$ tend to provide better coverage probability than their model-based counterparts for nonzero γ_1, especially when an uncorrelated but dependent important covariate is omitted from a Cox model (see Lin 1989).

Table 1. Empirical Size of the Model-Based and the Robust Statistics for Testing the Effect of Z_1 Under the Falsely Assumed Cox Model $\lambda(t; Z_1, Z_2) = \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2)$ at the .05 Nominal Level

	Model-based				Robust			
	Wald		Score		Wald		Score	
True model	n = 100	n = 50	n = 100	n = 50	n = 100	n = 50	n = 100	n = 50
1. $\lambda(t) = \exp(.2Z_2 + Z_3)$	.054	.056	.055	.055	.059	.075	.056	.064
2. $\lambda(t) = \exp(.2Z_2 + Z_1^2)$	.128	.137	.128	.139	.061	.069	.054	.057
3. $\lambda(t) = \exp(Z_1^2)$	.122	.112	.122	.114	.049	.058	.046	.050
4. $\lambda(t) = \exp(.2Z_2 + Z_1^2 + Z_3)$	.126	.130	.127	.132	.058	.068	.057	.057
5. $\lambda(t) = 1 + .5Z_2$	.042	.046	.043	.047	.051	.064	.048	.057
6. $\lambda(t) = 1 + .5Z_2 + Z_1^2$	.045	.039	.045	.039	.055	.060	.053	.050
7. $\lambda(t) = \log(2 + .5Z_2)$	.045	.045	.047	.050	.057	.066	.053	.064
8. $\lambda(t) = \log(2 + .5Z_2 + Z_1^2)$	.036	.039	.037	.040	.050	.066	.047	.054
9. $\log T = -.5Z_2 + \phi$	.078	.066	.078	.068	.075	.077	.069	.067
10. $\log T = -.5Z_2 - Z_1^2 + \phi$	.184	.185	.185	.188	.052	.069	.048	.063
11. $T = \exp(-.5Z_2) + \varepsilon$	.082	.091	.083	.094	.071	.092	.067	.081
12. $T = \exp(-.5Z_2 - Z_1^2) + \varepsilon$	.101	.106	.101	.108	.061	.082	.053	.067

NOTE: Z_1, Z_2, and Z_3 are independent standard normal variables truncated at ±5 in rows 1–4 and 9–12. Z_1 and Z_2 are independent standard normal variables truncated at ±1.96 in rows 5–8. ϕ is a zero-mean normal variable with .5 standard deviation. ε is a standard exponential variable. Each entry is based on 1,000 Monte Carlo replications without censoring. Uniform random numbers are generated through an algorithm provided by Press, Flannery, Teukolsky, and Vetterling (1986, pp. 196–197). Transformation and rejection methods are then applied to create random variables from other distributions. All calculations are programmed in FORTRAN-77 with double arithmetic precision.

APPENDIX: PROOF OF THEOREM 2.1

It is simple to show that the assumptions made in Section 2 imply that there exists a neighborhood $\mathcal{B}$ of β^* such that, for each $\tau < \infty$ and $r = 0, 1, 2$,

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \| S^{(r)}(\beta, t) - s^{(r)}(\beta, t)\| \to 0$$

in probability as $n \to \infty$ and $s^{(r)}(\beta, t)/s^{(0)}(\beta, t)$ ($r = 1, 2$) are bounded on $\mathcal{B} \times [0, \tau]$. We will use these facts and the assumptions repeatedly in the proof without referring to them explicitly.

Notice that

$$U(\beta) = \sum_{i=1}^{n} \int_0^\infty Z_i(t)\, dN_i(t) - \int_0^\infty \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}\, d\overline{N}(t),$$

where $N_i(t) = I\{X_i \le t, \delta_i = 1\}$ and $\overline{N}(t) = \sum N_i(t)$. Taylor expansion of $U(\hat\beta)$ around β^* results in $n^{1/2}(\hat\beta - \beta^*) = \hat A^{-1}(\tilde\beta)n^{-1/2}U(\beta^*)$, where $\tilde\beta$ is on the line segment between $\hat\beta$ and β^*. The consistency of $\hat A(\tilde\beta)$ for $A(\beta^*)$ can be easily established by the techniques used in the proofs of theorems 3.2 and 4.2 of Andersen and Gill (1982). We will show that $n^{-1/2}U(\beta^*)$ can be expressed as a sum of n iid random vectors plus terms that converge in probability to 0. First, let us rewrite $n^{-1/2}U(\beta^*)$ as

$$n^{-1/2} \sum_{i=1}^{n} \int_0^\infty Z_i(t)\, dN_i(t) - n^{1/2} \int_0^\infty \frac{s^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)}\, d\{\tilde F_n(t) - \tilde F(t)\} - n^{1/2} \int_0^\infty \frac{S^{(1)}(\beta^*, t)}{S^{(0)}(\beta^*, t)}\, d\tilde F(t) - n^{1/2} \int_0^\infty \left\{ \frac{S^{(1)}(\beta^*, t)}{S^{(0)}(\beta^*, t)} - \frac{s^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} \right\} d\{\tilde F_n(t) - \tilde F(t)\}, \quad (A.1)$$

where $\tilde F_n(t) = \overline{N}(t)/n$ and $\tilde F(t) = E\{\tilde F_n(t)\}$. Now, $n^{1/2}\{\tilde F_n(t) - \tilde F(t)\}$ converges in distribution to a zero-mean Gaussian process. Therefore, the last term in (A.1) is $o_p(1)$. The third term in (A.1) can be shown to be

$$n^{1/2} \int_0^\infty s^{(0)}(\beta^*, t)^{-1} \times \left[S^{(1)}(\beta^*, t) - \frac{s^{(1)}(\beta^*, t)}{s^{(0)}(\beta^*, t)} \{S^{(0)}(\beta^*, t) - s^{(0)}(\beta^*, t)\} \right] \times d\tilde F(t) + o_p(1).$$

Combining the foregoing expression with the first two terms in (A.1), we observe that $n^{-1/2}U(\beta^*)$ is asymptotically equivalent to $n^{-1/2} \sum w_i(\beta^*)$, where

$$w_i(\beta) = \int_0^\infty \left\{ Z_i(t) - \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right\} dN_i(t) - \int_0^\infty \frac{Y_i(t)\exp\{\beta' Z_i(t)\}}{s^{(0)}(\beta, t)} \left\{ Z_i(t) - \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right\} d\tilde F(t).$$

Notice that $w_i(\beta^*)$ ($i = 1, \ldots, n$) are iid. Thus the asymptotic normality of $n^{-1/2}U(\beta^*)$ follows from the multivariate central limit theorem. The corresponding asymptotic covariance matrix is $B(\beta^*) = E\{w_1(\beta^*)^{\otimes 2}\}$. Therefore, $n^{1/2}(\hat\beta - \beta^*)$ is asymptotically normal with zero mean and covariance matrix $V(\beta^*) = A^{-1}(\beta^*)B(\beta^*)A^{-1}(\beta^*)$. The matrix $B(\beta^*)$ can be estimated by $\hat B(\hat\beta)$, where $\hat B(\beta) = n^{-1} \sum W_i(\beta)^{\otimes 2}$. Note that $W_i(\beta)$ is obtained from $w_i(\beta)$ by replacing $s^{(0)}(\beta, t)$, $s^{(1)}(\beta, t)$, and $\tilde F(t)$ by $S^{(0)}(\beta, t)$, $S^{(1)}(\beta, t)$ and $\tilde F_n(t)$, respectively. The consistency of $\hat B(\hat\beta)$ for $B(\beta^*)$ can be established by the properties of the empirical distribution function and some elementary probability arguments. It follows that $V(\beta^*)$ can be consistently estimated by $\hat V(\hat\beta)$, which completes the proof.

[*Received August 1988. Revised March 1989.*]

REFERENCES

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Bailey, K. R. (1983), "The Asymptotic Joint Distribution of Regression and Survival Parameter Estimates in the Cox Regression Model," *The Annals of Statistics*, 11, 39–48.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

——— (1975), "Partial Likelihood," *Biometrika*, 62, 269–276.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, New York: Chapman & Hall.

Gail, M. H., Tan, W. Y., and Piantadosi, S. (1988), "Tests for No Treatment Effect in Randomized Clinical Trials," *Biometrika*, 75, 57–64.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984), "Biased Estimates of Treatment Effect in Randomized Experiments With Nonlinear Regressions and Omitted Covariates," *Biometrika*, 71, 431–444.

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates

Under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability,* Berkeley: University of California Press, pp. 221–233.

Kent, J. T. (1982), "Robust Properties of Likelihood Ratio Tests," *Biometrika,* 69, 19–27.

Lagakos, S. W. (1988), "The Loss in Efficiency From Misspecifying Covariates in Proportional Hazards Regression Models," *Biometrika,* 75, 156–160.

Lagakos, S. W., and Schoenfeld, D. A. (1984), "Properties of Proportional-Hazards Score Tests Under Misspecified Regression Models," *Biometrics,* 40, 1037–1048.

Lin, D. Y. (1989), "Goodness-of-Fit Tests and Robust Statistical Inference for the Cox Proportional Hazards Model," unpublished Ph.D. thesis, University of Michigan, Dept. of Biostatistics.

Morgan, T. M. (1986), "Omitting Covariates From the Proportional Hazards Model," *Biometrics,* 42, 993–995.

Næs, T. (1982), "The Asymptotic Distribution of the Estimator for the Regression Parameter in Cox's Regression Model," *Scandinavian Journal of Statistics,* 9, 107–115.

O'Neill, T. J. (1986), "Inconsistency of the Misspecified Proportional Hazards Model," *Statistics & Probability Letters,* 4, 219–222.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986), *Numerical Recipes: The Art of Scientific Computing,* Cambridge, U.K.: Cambridge University Press.

Royall, R. M. (1986), "Model Robust Confidence Intervals Using Maximum Likelihood Estimators," *International Statistical Review,* 54, 221–226.

Solomon, P. J. (1984), "Effect of Misspecification of Regression Models in the Analysis of Survival Data," *Biometrika,* 71, 291–298; Amendment (1986), 73, 245.

Struthers, C. A., and Kalbfleisch, J. D. (1986), "Misspecified Proportional Hazard Models," *Biometrika,* 73, 363–369.

Tsiatis, A. A. (1981), "A Large Sample Study of Cox's Regression Model," *The Annals of Statistics,* 9, 93–108.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica,* 50, 1–25.