Semiparametric Analysis of the Additive Risk Model

D. Y. Lin, Zhiliang Ying

Biometrika, Volume 81, Issue 1 (Mar., 1994), 61-71.

Biometrika (1994), **81**, 1, *pp*. 61–71
Printed in Great Britain

Semiparametric analysis of the additive risk model

By D. Y. LIN
Department of Biostatistics, SC-32, University of Washington, Seattle, Washington 98195, U.S.A.

AND ZHILIANG YING
Department of Statistics, University of Illinois, Champaign, Illinois 61820, U.S.A.

SUMMARY

In contrast to the proportional hazards model, the additive risk model specifies that the hazard function associated with a set of possibly time-varying covariates is the sum of, rather than the product of, the baseline hazard function and the regression function of covariates. This formulation describes a different aspect of the association between covariates and the failure time than the proportional hazards model, and is more plausible than the latter for many applications. In the present paper, simple procedures with high efficiencies are developed for making inference about the regression parameters under the additive risk model with an unspecified baseline hazard function. The subject-specific survival estimation is also studied. The proposed techniques resemble the partial-likelihood-based methods for the proportional hazards model. A real example is provided.

Some key words: Adaptive estimation; Censoring; Counting process; Excess risk; Failure time; Information bound; Martingale; Partial likelihood; Proportional hazards; Regression; Survival data; Time-dependent covariate; Truncation.

1. INTRODUCTION

The additive and multiplicative risk models provide the two principal frameworks for studying the association between risk factors and disease occurrence or death. As elucidated by Breslow & Day (1980, pp. 53–9; 1987, pp. 122–31), both modelling approaches have sound biological and empirical bases, providing complementary information about the association. The hazard function for the failure time T associated with a p-vector of possibly time-varying covariates $Z(.)$ takes the form

$$\lambda(t; Z) = \lambda_0(t) + \beta_0' Z(t) \tag{1·1}$$

under the additive risk model (Cox & Oakes, 1984, p. 74; Thomas, 1986; Breslow & Day, 1987, p. 182) and takes the form

$$\lambda(t; Z) = \lambda_0(t) e^{\gamma_0' Z(t)} \tag{1·2}$$

under the multiplicative risk model (Cox, 1972), where β_0 and γ_0 are p-vectors of regression parameters. The additive and multiplicative models intersect when $\lambda_0(.)$ is time-invariant and the exponential regression form in (1·2) is replaced by the linear form $\{1 + \gamma_0' Z(t)\}$, in which case $\beta_0 = \lambda_0 \gamma_0$.

In biomedical applications, the failure time T is often subject to left-truncation and right-censoring. Furthermore, due to the complexity of biological processes, it is desirable not to parameterize the baseline hazard function $\lambda_0(.)$. The main statistical challenge then becomes the semiparametric estimation of regression parameters with left-truncated and right-censored observations. The estimation of the baseline hazard function and subject-specific survival curves may also be of scientific interest.

In order to draw semiparametric inference for model (1·2), Cox (1972; 1975) introduced the partial likelihood approach, which eliminates the nuisance quantity $\lambda_0(.)$ from the score function for γ_0. The resulting maximum partial likelihood estimator possesses asymptotic properties similar to those of the standard maximum likelihood estimator (Tsiatis, 1981; Andersen & Gill, 1982). Such desirable theoretical properties, together with the simple interpretation of the results and the wide availability of computer programs, have made the multiplicative risk model the current method of choice in survival analysis.

Although additive risk models in various forms have been eloquently advocated and successfully utilized by numerous authors, e.g. Aalen (1980), Breslow & Day (1980, pp. 55, 58), Pocock, Gore & Kerr (1982), Buckley (1984), Pierce & Preston (1984), Thomas (1986), Breslow & Day (1987, pp. 122–31, 142–46), Aalen (1989), Huffer & McKeague (1991), no satisfactory semiparametric methods of estimation have been developed for model (1·1). The lack of progress is attributed to the fact that the partial likelihood approach cannot be directly used to eliminate the nuisance function $\lambda_0(.)$ in estimating β_0.

In the next section of this paper, a simple semiparametric estimating function for β_0 is constructed, which mimics the martingale feature of the partial likelihood score function for γ_0. The resulting estimator, which takes an explicit form, is consistent and asymptotically normal with an easily estimated covariance matrix. Also presented in § 2 is an estimator for the cumulative baseline hazard function under model (1·1), which parallels the Breslow (1972) estimator for the corresponding quantity under model (1·2). A real example is provided in § 3 for illustration. In § 4, semiparametric efficiencies of the proposed estimator and some alternatives are studied. Several remarks follow in § 5.

2. INFERENCE PROCEDURES

Consider a set of n independent subjects such that the counting process $\{N_i(t); t \geqslant 0\}$ for the ith subject in the set records the number of observed events up to time t. The intensity function for $N_i(t)$ is given by

$$Y_i(t)\, d\Lambda(t; Z_i) = Y_i(t)\{d\Lambda_0(t) + \beta_0' Z_i(t)\, dt\} \tag{2·1}$$

under model (1·1), and by

$$Y_i(t)\, d\Lambda(t; Z_i) = Y_i(t) e^{\gamma_0' Z_i(t)}\, d\Lambda_0(t) \tag{2·2}$$

under model (1·2), where $Y_i(t)$ is a 0–1 predictable process indicating, by the value 1, whether the ith subject is at risk at time t, $Z_i(.)$ is the covariate process for the ith subject, and

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

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

$$N_i(t) = M_i(t) + \int_0^t Y_i(u)\, d\Lambda(u; Z_i), \tag{2·3}$$

where $M_i(.)$ is a local square integrable martingale (Andersen & Gill, 1982). In view of relationship (2·3), it is natural to estimate $\Lambda_0(t)$ by

$$\hat{\Lambda}_0(\hat{\beta}, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\hat{\beta}' Z_i(u)\, du\}}{\sum_{j=1}^n Y_j(u)} \tag{2·4}$$

under model (2·1), and by

$$\tilde{\Lambda}_0(\hat{\gamma}, t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{j=1}^n Y_j(u) e^{\hat{\gamma}' Z_j(u)}} \tag{2·5}$$

under model (2·2), where $\hat{\beta}$ and $\hat{\gamma}$ are consistent estimators. Estimator (2·5) is commonly credited to Breslow (1972).

The partial likelihood score function for γ_0 can be written as

$$\sum_{i=1}^n \int_0^\infty Z_i(t)\{dN_i(t) - Y_i(t) e^{\gamma' Z_i(t)}\, d\tilde{\Lambda}_0(\gamma, t)\}. \tag{2·6}$$

Mimicking (2·6), we propose to estimate β_0 from the following estimating function

$$U(\beta) = \sum_{i=1}^n \int_0^\infty Z_i(t)\{dN_i(t) - Y_i(t)\, d\hat{\Lambda}_0(\beta, t) - Y_i(t)\beta' Z_i(t)\, dt\},$$

which is equivalent to

$$U(\beta) = \sum_{i=1}^n \int_0^\infty \{Z_i(t) - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)\beta' Z_i(t)\, dt\}, \tag{2·7}$$

where

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

The resulting estimator takes the explicit form

$$\hat{\beta} = \left[\sum_{i=1}^n \int_0^\infty Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2}\, dt\right]^{-1} \left[\sum_{i=1}^n \int_0^\infty \{Z_i(t) - \bar{Z}(t)\}\, dN_i(t)\right], \tag{2·8}$$

where $a^{\otimes 2} = aa'$.

A simple algebraic manipulation yields

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

which is a martingale integral. It then follows from standard counting process arguments (Andersen & Gill, 1982) that the random vector $n^{-\frac{1}{2}} U(\beta_0)$ converges weakly to a p-variate normal with mean zero and with a covariance matrix which can be consistently estimated by

$$B = n^{-1} \sum_{i=1}^n \int_0^\infty \{Z_i(t) - \bar{Z}(t)\}^{\otimes 2}\, dN_i(t).$$

Furthermore, the random vector $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converges weakly to a p-variate normal with mean zero and with a covariance matrix which can be consistently estimated by

$A^{-1}BA^{-1}$, where

$$A = n^{-1} \sum_{i=1}^{n} \int_0^{\infty} Y_i(t)\{Z_i(t) - \bar{Z}(t)\}^{\otimes 2}\, dt.$$

Calculations of $\hat{\beta}$, A and B are all straightforward, especially for time-independent covariates. They may even be calculated by hand for small data sets. A general FORTRAN program is available from the first author. Simulation studies have shown that the aforementioned asymptotic approximations are very satisfactory for practical sample sizes; the degrees of accuracy are fairly comparable to those of the partial likelihood procedures under model (1·2).

The estimator for $\Lambda_0(.)$ given in (2·4) provides the basis for estimating/predicting survival experience. Standard counting process techniques can again be employed to prove that the process $n^{\frac{1}{2}}\{\hat{\Lambda}_0(\hat{\beta}, .) - \Lambda_0(.)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t, s) $(t \geqslant s)$ can be consistently estimated by

$$\int_0^s \frac{n\sum_{i=1}^n dN_i(u)}{\{\sum_{j=1}^n Y_j(u)\}^2} + C'(t)A^{-1}BA^{-1}C(s) - C'(t)A^{-1}D(s) - C'(s)A^{-1}D(t),$$

where

$$C(t) = \int_0^t \bar{Z}(u)\, du, \quad D(t) = \int_0^t \frac{\sum_{i=1}^n \{Z_i(u) - \bar{Z}(u)\}\, dN_i(u)}{\sum_{j=1}^n Y_j(u)}.$$

Let $S(t; z)$ denote the survival function for an individual with a given covariate vector $z(.)$. Then it is natural to estimate $S(t; z)$ by

$$\hat{S}(t; z) = \exp\left\{-\hat{\Lambda}_0(\hat{\beta}, t) - \int_0^t \hat{\beta}' z(u)\, du\right\}. \qquad (2·9)$$

The process $n^{\frac{1}{2}}\{\hat{S}(.; z) - S(.; z)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t, s) $(t \geqslant s)$ can be consistently estimated by

$$\hat{S}(t; z)\hat{S}(s; z)\left[\int_0^s \frac{n\sum_{i=1}^n dN_i(u)}{\{\sum_{j=1}^n Y_j(u)\}^2} + G'(t; z)A^{-1}BA^{-1}G(s; z) \right.$$
$$\left. + G'(t; z)A^{-1}D(s) + G'(s; z)A^{-1}D(t)\right],$$

where

$$G(t; z) = \int_0^t \{z(u) - \bar{Z}(u)\}\, du.$$

The survival function estimation has also been implemented in our computer program.

As they stand, estimators (2·4) and (2·9) may not always be monotone in t. However, simple modifications can be made to ensure monotonicity while preserving the given asymptotic properties. For example, we may define

$$\hat{\Lambda}_0^*(t) = \max_{s \leqslant t} \hat{\Lambda}_0(\hat{\beta}, s), \quad \hat{S}^*(t; z) = \min_{s \leqslant t} \hat{S}(s; z).$$

Under appropriate regularity conditions, $\hat{\Lambda}_0^*$ and $\hat{\Lambda}_0$ are asymptotically equivalent in the sense that $\hat{\Lambda}_0^* - \hat{\Lambda}_0 = o_p(n^{-\frac{1}{2}})$ so that $n^{\frac{1}{2}}(\hat{\Lambda}_0^* - \Lambda_0)$ converges to the same limiting distribution as $n^{\frac{1}{2}}(\hat{\Lambda}_0 - \Lambda_0)$. Here we only provide a heuristic argument for this equivalence. Suppose that $\lambda_0(t)$ is positive for every t. Then for $n^{-2/3} \leqslant \eta \leqslant n^{-1/3}$,

$$\hat{\Lambda}_0(t) - \hat{\Lambda}_0(t-\eta) = \lambda_0(t^\dagger)\eta + [\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} - \{\hat{\Lambda}_0(t-\eta) - \Lambda_0(t-\eta)\}],$$

where $t^\dagger \in (t-\eta, t)$ and the argument $\hat{\beta}$ in $\hat{\Lambda}_0$ is suppressed. By the asymptotic linearity of $\hat{\Lambda}_0$, the term inside the square brackets of the preceding equation can be shown to be of the order $o_p\{(\eta/n)^{\frac{1}{2}-\varepsilon}\} = o_p(\eta^{1+\varepsilon})$ for some $\varepsilon > 0$, implying that $\hat{\Lambda}_0(t) \geqslant \hat{\Lambda}_0(t-\eta)$ for large n. Moreover, uniformly in $s \leqslant t - n^{-1/3}$,

$$\hat{\Lambda}_0(t) - \hat{\Lambda}_0(s) \geqslant \lambda_0(t^*)n^{-1/3} + O_p(n^{-1/2}) > 0,$$

where $t^* \in (s, t)$. Combining the case of $s \leqslant t - n^{-1/3}$ with that of $s \in [t - n^{-1/3}, t - n^{-2/3}]$, we have

$$\hat{\Lambda}_0(t - n^{-2/3}) \leqslant \hat{\Lambda}_0^*(t - n^{-2/3}) \leqslant \hat{\Lambda}_0(t)$$

for large n. Hence, $\hat{\Lambda}_0^*$ and $\hat{\Lambda}_0$ are equivalent since $\hat{\Lambda}_0(t) - \hat{\Lambda}_0(t - n^{-2/3}) = o_p(n^{-1/2})$. The equivalence between $\hat{S}^*$ and $\hat{S}$ can be argued in a similar manner.

3. A real example

We now apply the methods described in the last section to the South Wales nickel refiners study (Breslow & Day, 1987, Appendix ID). Men employed in a nickel refinery in South Wales were investigated to determine the risk of developing carcinoma of the bronchi and nasal sinuses associated with the refining of nickel. The cohort was identified using the weekly paysheets of the company and followed from the year 1934 until 1981. Appendix VIII of Breslow & Day (1987) contained complete records for 679 workers employed before 1925, to whom attention is henceforth confined. The follow-up through 1981 uncovered 137 lung cancer deaths among men aged 40–85 years and 56 deaths from cancer of the nasal sinus. Since the workers had been working in the company for various periods of time before the follow-up was initiated, their survival times were subject to left truncation. A right-censored observation arose either because the worker died from a competing cause or because he was still alive on the date of data listings.

Breslow & Day (1987, § 4.10) fitted both relative and excess risk models to grouped data on lung cancer deaths among the Welsh nickel refiners, and concluded that the two models fitted the data equally well. These authors also analyzed the continuous data on the nasal sinus cancer mortality using the proportional hazards model (Breslow & Day, 1987, pp. 222–3). They considered the survival time to be years since first employment and found three significant risk factors: age at first employment, AFE, year at first employment, YFE, and exposure level, EXP. Their final results are given under the multiplicative risk model columns in Table 1.

The additive risk model columns in Table 1 display the results from fitting (1·1) to the same data. It is interesting to note that the parameter estimates under model (1·1) are much smaller than those of model (1·2), which is not surprising since the former pertain to the risk differences whereas the latter to the risk ratios. The chi-squared statistics for testing individual covariate effects, however, are very comparable between the two models. Incidentally, in his analysis of a clinical trial, Aalen (1989) also found close agreement between test statistics for individual covariate effects when using additive and multiplicative risk models.

Table 1. *Multiplicative and additive risk analyses of time from the first employment to the nasal sinus cancer death for the Welsh nickel refiners study*

Parameters	Multiplicative risk model	Additive risk model	Parameters	Multiplicative risk model	Additive risk model
log (AFE−10)			(YFE−1915)²/100		
Est.	2·22	0·00431	Est.	−1·26	−0·00496
SE	0·44	0·00083	SE	0·51	0·00209
Est./SE	5·09	5·16	Est./SE	−2·48	−2·37
P-value	<0·00001	<0·00001	*P*-value	0·013	0·018
(YFE−1915)/10			log (EXP + 1)		
Est.	−0·09	0·00005	Est.	0·77	0·00373
SE	0·32	0·00102	SE	0·17	0·00093
Est./SE	−0·30	0·05	Est./SE	4·40	4·01
P-value	0·76	0·96	*P*-value	0·00001	0·00006

AFE, age at first employment; YFE, year at first employment; EXP, exposure level.

Survival estimates can be quite different between the additive and multiplicative risk models. Figure 1 shows the estimates of the survival curve for a worker with AFE $= 25$, YFE $= 1915$ and EXP $= 1$ under the two models. The selected covariate values are roughly the sample medians. In this case, the estimate based on the multiplicative risk model is considerably higher than that of the additive risk model. Note also that the former is far more discrete than the latter.

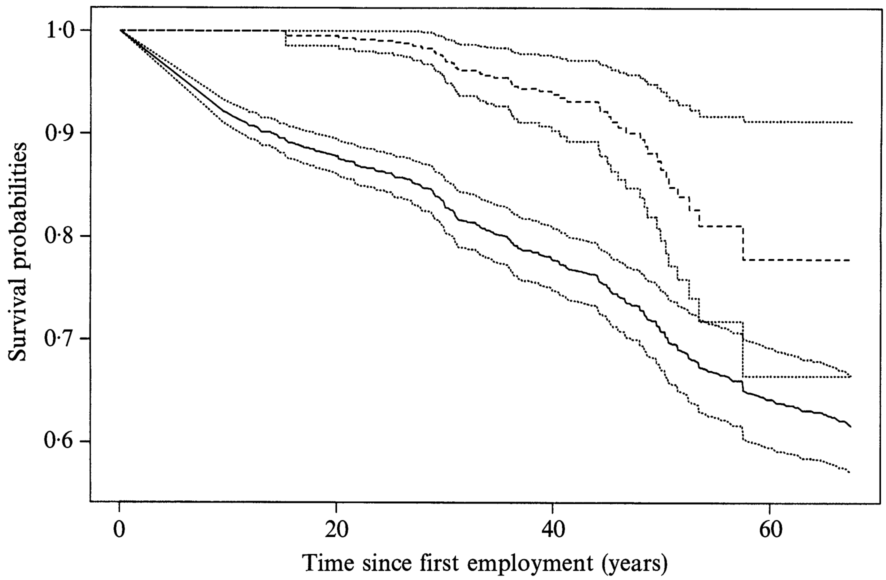

Fig. 1. Estimates of the survival function for a nickel refiner with AFE $= 25$, YEF $=$ 1915 and EXP $= 1$ under the additive and multiplicative risk models, shown by the solid and dashed curves, respectively, along with the pointwise 95% confidence limits, shown by dotted curves.

4. Efficiency considerations

Since estimating function (2·7) was introduced in a somewhat ad hoc fashion, a question naturally arises as to how efficient the resulting inference procedures are. Here we provide insights into this problem by examining the semiparametric information bound.

Following Lai & Ying (1992), we consider the family of parametric sub-models

$$\lambda(t; Z) = \lambda_0(t) + \theta'\mu(t) + \beta'Z(t), \tag{4·1}$$

where θ and β are p-vectors of unknown parameters, and $\lambda_0(.)$ and $\mu(.)$ are fixed functions. As explained by Bickel et al. (1993), finding the semiparametric information bound for β at β_0 is tantamount to finding the supremum parametric information bound for β in (4·1) at $\beta = \beta_0$ and $\theta = 0$ among all choices of $\mu(.)$.

The log-likelihood function for (4·1) is

$$l(\beta, \theta) = \sum_{i=1}^{n} \left[\int_0^\infty \log\{\lambda_0(t) + \theta'\mu(t) + \beta'Z_i(t)\}\, dN_i(t) \right.$$
$$\left. - \int_0^\infty Y_i(t)\{\lambda_0(t) + \theta'\mu(t) + \beta'Z_i(t)\}\, dt \right].$$

Denote the limiting Fisher information matrix at $\beta = \beta_0$ and $\theta = 0$ by

$$\begin{pmatrix} I_{\beta\beta}(\mu) & I_{\beta\theta}(\mu) \\ I'_{\beta\theta}(\mu) & I_{\theta\theta}(\mu) \end{pmatrix}.$$

Then

$$I_{\beta\beta}(\mu) = -\lim_{n\to\infty} n^{-1} E \left.\frac{\partial^2 l(\beta, \theta)}{\partial\beta^2}\right|_{\beta=\beta_0, \theta=0} = \lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^{n} E\left\{\frac{Y_i(t)Z_i(t)^{\otimes 2}}{\lambda_0(t) + \beta_0'Z_i(t)}\right\} dt,$$

$$I_{\beta\theta}(\mu) = -\lim_{n\to\infty} n^{-1} E \left.\frac{\partial^2 l(\beta, \theta)}{\partial\beta\partial\theta}\right|_{\beta=\beta_0, \theta=0} = \lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^{n} E\left\{\frac{Y_i(t)Z_i(t)\mu(t)'}{\lambda_0(t) + \beta_0'Z_i(t)}\right\} dt,$$

$$I_{\theta\theta}(\mu) = -\lim_{n\to\infty} n^{-1} E \left.\frac{\partial^2 l(\beta, \theta)}{\partial\theta^2}\right|_{\beta=\beta_0, \theta=0} = \lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^{n} E\left\{\frac{Y_i(t)\mu(t)^{\otimes 2}}{\lambda_0(t) + \beta_0'Z_i(t)}\right\} dt,$$

where E denotes expectation.

The Cramér–Rao inequality entails that, for any regular semiparametric estimator $\hat\beta$ with $n^{\frac{1}{2}}(\hat\beta - \beta_0)$ converging to a zero-mean normal with covariance matrix Ω,

$$\Omega \geqslant \{I_{\beta\beta}(\mu) - I_{\beta\theta}(\mu) I_{\theta\theta}^{-1}(\mu) I'_{\beta\theta}(\mu)\}^{-1}$$

for every μ, where $Q_1 \geqslant Q_2$ means that $Q_1 - Q_2$ is nonnegative definite. The right-hand side of the preceding inequality reaches its maximum at $\mu(t) = \mu_0(t)$, where

$$\mu_0(t) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} E[Y_i(t)Z_i(t)/\{\lambda_0(t) + \beta_0'Z_i(t)\}]}{\sum_{i=1}^{n} E[Y_i(t)/\{\lambda_0(t) + \beta_0'Z_i(t)\}]},$$

which gives the information bound

$$\left(\lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^{n} E\left[\frac{Y_i(t)\{Z_i(t) - \mu_0(t)\}^{\otimes 2}}{\lambda_0(t) + \beta_0'Z_i(t)}\right] dt\right)^{-1}. \tag{4·2}$$

Therefore, an optimal estimating function for β_0 would be

$$U_{\text{opt}}(\beta) = \sum_{i=1}^{n} \int_0^\infty \frac{\{Z_i(t) - \tilde Z(t; \lambda_0, \beta_0)\}}{\lambda_0(t) + \beta_0'Z_i(t)} \{dN_i(t) - Y_i(t)\beta'Z_i(t)\, dt\}, \tag{4·3}$$

where

$$\tilde{Z}(t; \lambda_0, \beta) = \frac{\sum_{i=1}^n Y_i(t) Z_i(t)/\{\lambda_0(t) + \beta' Z_i(t)\}}{\sum_{i=1}^n Y_i(t)/\{\lambda_0(t) + \beta' Z_i(t)\}}.$$

It is straightforward to show that the limiting covariance matrix for the resulting estimator equals (4·2).

In view of (4·3), the estimating function (2·7) will be optimal if $\lambda_0(.)$ is constant and $\beta_0 = 0$, and should have high efficiencies for small β_0 and approximately constant $\lambda_0(.)$. The following examples indicate that the loss in efficiency may be small for nonzero β_0 and time-varying $\lambda_0(.)$.

Example 4·1. The first special case assumes $\beta_0 = 0$ and no censorship or truncation. In this case, $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converges to a zero-mean normal with covariance matrix $[\int \{1 - F_0(t)\} \, dt]^{-2} V^{-1}$, where the integral is over the range $(0, \infty)$ and where $F_0(.)$ is the distribution function corresponding to $\lambda_0(.)$ and V is the covariance matrix of Z. By comparing this covariance matrix with the information bound (4·2), we see that the relative efficiency is

$$\left[\int_0^\infty \{1 - F_0(t)\} \, dt\right]^2 \Big/ \int_0^\infty \{1 - F_0(t)\} \lambda_0^{-1}(t) \, dt.$$

If $\lambda_0(.)$ is time-invariant, then the efficiency is 1, implying that the proposed estimator is semiparametrically efficient. If $\lambda_0(.)$ is half-logistic, that is, $1 - F_0(t) = 2/(1 + e^t)$, then the efficiency is $(2 \log 2)^2/2 \simeq 0{\cdot}9609$.

Example 4·2. Next we consider the two-sample problem with nonzero β_0 in which $Z_i = 0$ for $i \leqslant n/2$ and $Z_i = 1$ for $i > n/2$. For simplicity, assume $\lambda_0(.) \equiv 1$ and no censorship or truncation. Then the relative efficiency can be evaluated using numerical integration. For $\beta_0 = 0{\cdot}5$, 1 and 1·5, the relative efficiencies are found to be 0·999, 0·996 and 0·993, respectively.

Adaptive estimators for β_0 may be constructed which achieve the semiparametric efficiency bound (4·2). Let us divide the entire sample into two disjoint subsets, the first of which contains the first n_1 subjects, where n_1 is the largest integer $\leqslant n/2$. Let $\hat{\beta}^{(1)}$ and $\hat{\lambda}_0^{(1)}(.)$ be some preliminary estimators for β_0 and $\lambda_0(.)$ calculated from the first subsample. This can be done by using (2·8) and by smoothing (2·4) on the basis of the first subsample. Similarly, $\hat{\beta}^{(2)}$ and $\hat{\lambda}_0^{(2)}(.)$ are obtained from the second subsample. We then estimate β_0 by the estimating function

$$U_{\mathrm{adp}}(\beta) = \sum_{i=1}^{n_1} \int_0^\infty \frac{\{Z_i(t) - \tilde{Z}^{(1)}(t; \hat{\lambda}_0^{(2)}, \hat{\beta}^{(2)})\}}{\hat{\lambda}_0^{(2)}(t) + \hat{\beta}^{(2)\prime} Z_i(t)} \{dN_i(t) - Y_i(t) \beta' Z_i(t) \, dt\}$$
$$+ \sum_{i=n_1+1}^{n} \int_0^\infty \frac{\{Z_i(t) - \tilde{Z}^{(2)}(t; \hat{\lambda}_0^{(1)}, \hat{\beta}^{(1)})\}}{\hat{\lambda}_0^{(1)}(t) + \hat{\beta}^{(1)\prime} Z_i(t)} \{dN_i(t) - Y_i(t) \beta' Z_i(t) \, dt\},$$

where $\tilde{Z}^{(1)}(t; \lambda_0, \beta)$ and $\tilde{Z}^{(2)}(t; \lambda_0, \beta)$ are the same as $\tilde{Z}(t; \lambda_0, \beta)$ except that the summations are taken from 1 to n_1 and from $(n_1 + 1)$ to n, respectively. The resulting estimator, denoted by $\hat{\beta}_{\mathrm{adp}}$, takes an explicit form similar to (2·8). Clearly,

$$U_{\mathrm{adp}}(\beta_0)=\sum_{i=1}^{n_1}\int_0^\infty \frac{\{Z_i(t)-\tilde{Z}^{(1)}(t;\hat{\lambda}_0^{(2)},\hat{\beta}^{(2)})\}}{\hat{\lambda}_0^{(2)}(t)+\hat{\beta}^{(2)\prime}Z_i(t)}\,dM_i(t)$$
$$+\sum_{i=n_1+1}^{n}\int_0^\infty \frac{\{Z_i(t)-\tilde{Z}^{(2)}(t;\hat{\lambda}_0^{(1)},\hat{\beta}^{(1)})\}}{\hat{\lambda}_0^{(1)}(t)+\hat{\beta}^{(1)\prime}Z_i(t)}\,dM_i(t). \tag{4·4}$$

The first term on the right-hand side of (4·4) is a sum of integrals of predictable processes with respect to the martingales $\{M_i(t); i=1,\ldots,n_1\}$, where the σ-filtration $\mathscr{F}_t^{(1)}$ is generated by

$$\{N_i(s), Y_i(s), Z_i(s); s\leqslant t, 1\leqslant i\leqslant n_1\},\quad \{N_j(u), Y_j(u), Z_j(u); 0\leqslant u<\infty, n_1<j\leqslant n\}.$$

Standard counting process arguments can then be used to show that this term is asymptotically equivalent to the same expression but with estimators $\hat{\lambda}_0^{(2)}$ and $\hat{\beta}^{(2)}$ replaced by their true values λ_0 and β_0. Likewise, we can replace $\hat{\lambda}_0^{(1)}$ and $\hat{\beta}^{(1)}$ by λ_0 and β_0 in the second term on the right-hand side of (4·4) without affecting its asymptotic behaviour. Hence, $U_{\mathrm{adp}}(\beta_0)$ is asymptotically equivalent to $U_{\mathrm{opt}}(\beta_0)$, which entails the optimality of $\hat{\beta}_{\mathrm{adp}}$.

The aforementioned adaptive procedure may not be heartily advocated for small samples since it is difficult to estimate $\lambda_0(.)$ well. We now provide a compromise between $U(\beta)$ and $U_{\mathrm{adp}}(\beta)$, which was suggested by J. Huang. Suppose that $\lambda_0^*(.)$ and β_0^* are guesses of $\lambda_0(.)$ and β_0 based on prior knowledge. Then β_0 may be estimated by the following estimating function

$$U^*(\beta)=\sum_{i=1}^{n}\int_0^\infty \frac{\{Z_i(t)-\tilde{Z}(t;\lambda_0^*,\beta_0^*)\}}{\lambda_0^*(t)+\beta_0^{*\prime}Z_i(t)}\{dN_i(t)-Y_i(t)\beta' Z_i(t)\,dt\}.$$

The asymptotic distributional properties for the resulting estimator $\hat{\beta}^*$ follow immediately from the fact that

$$U^*(\beta_0)=\sum_{i=1}^{n}\int_0^\infty \frac{\{Z_i(t)-\tilde{Z}(t;\lambda_0^*,\beta_0^*)\}}{\lambda_0^*(t)+\beta_0^{*\prime}Z_i(t)}\,dM_i(t).$$

Obviously, $\hat{\beta}^*$ will have high efficiency if λ_0^* and β_0^* are close to their true values λ_0 and β_0. One may also replace β_0^* in $U^*(\beta)$ by β at the expense of a more complicated estimating procedure.

5. Remarks

The current work makes the additive risk model a practical alternative to the proportional hazards model. The choice between the two models will normally be an empirical matter. Although in theory either model can provide adequate fit to a given data set if appropriate time-dependent covariates are introduced, the more parsimonious one will undoubtedly be more appealing to medical investigators. It seems desirable to fit both models to the same data set as they inform us about two quite different aspects of the association between risk factors and disease or death.

Model (1·1) assumes the linear regression form $\beta_0' Z(t)$, which has an easy interpretation and leads to exceedingly simple inference procedures. A limitation of this representation is that $\beta_0' Z(t)$ needs to be constrained so that the right-hand side is nonnegative. A similar constraint is needed for the proportional hazards model if the exponential regression function in (1·2) is replaced by the linear form $\{1+\gamma_0' Z(t)\}$. One may avoid the constraint for the additive risk model by substituting $e^{\beta_0' Z(t)}$ for $\beta_0' Z(t)$, in which case $\lambda_0(t)$ corresponds to the hazard function under $\beta_0' Z(t) = -\infty$ rather than under $Z(t) = 0$. The ideas presented in §§ 2 and 4 can be applied to the general regression function $g\{\beta_0' Z(t)\}$, and the basic conclusions continue to hold. The resulting procedures, however, may not enjoy some of the good properties of the linear form. For example, there is in general no explicit solution to the estimating equation, and the Newton–Raphson algorithm will be required. Furthermore, the derivative matrix of the estimating function is not necessarily positive definite, which makes the analysis more complicated both numerically and theoretically. In the University of Washington, Department of Biostatistics Technical Report No. 129, we provide a rigorous asymptotic theory for general additive-multiplicate intensity models.

Aalen (1980; 1989), Huffer & McKeague (1991) and Andersen et al. (1993) studied the following nonparametric additive risk model

$$\lambda(t; Z) = \sum_{j=1}^{p} Z_j(t)\alpha_j(t). \tag{5·1}$$

These authors provided estimators for the cumulative regression functions

$$A_j(t) = \int_0^t \alpha_j(u)\, du \quad (j = 1, \dots, p)$$

based on the least-squares type methods. Recently, McKeague (1992) suggested a more restrictive version of model (5·1), $\lambda(t; Z) = \alpha(t)' X(t) + \beta_0' W(t)$, where Z is partitioned into two parts X and W. He also indicated how one might analyze this model.

Another useful model, which was motivated by the aforementioned McKeague model, specifies that

$$\lambda(t; Z) = \lambda_0(t) e^{\gamma_0' X(t)} + \beta_0' W(t).$$

Obviously, this model includes both models (1·1) and (1·2) as special cases. By the rationale given in § 2, a natural estimating function for $\theta_0 = (\gamma_0', \beta_0')'$ is

$$U(\theta) = \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i(t) - \frac{\sum_{j=1}^n Y_j(t) e^{\gamma' X_j(t)} Z_j(t)}{\sum_{j=1}^n Y_j(t) e^{\gamma' X_j(t)}} \right\} \{dN_i(t) - Y_i(t)\beta' W_i(t)\, dt\},$$

the resulting estimator being denoted by $\hat\theta = (\hat\gamma', \hat\beta')'$. It can be shown that $n^{\frac{1}{2}}(\hat\theta - \theta_0)$ is asymptotically zero-mean normal with a covariance matrix which can be consistently estimated by $\{A(\hat\theta)\}^{-1} B(\hat\gamma) \{A(\hat\theta)'\}^{-1}$, where $A(\theta) = -n^{-1} \partial U(\theta)/\partial\theta$, and

$$B(\gamma) = n^{-1} \sum_{i=1}^{n} \int_0^\infty \left\{ Z_i(t) - \frac{\sum_{j=1}^n Y_j(t) e^{\gamma' X_j(t)} Z_j(t)}{\sum_{j=1}^n Y_j(t) e^{\gamma' X_j(t)}} \right\}^{\otimes 2} dN_i(t).$$

There are a number of important issues to be addressed for model (1·1). We are currently investigating the following topics: (i) generalizing estimating function (2·7) to the case of multivariate failure time data, (ii) developing methods for checking the adequacy of model (1·1) and for discriminating between models (1·1) and (1·2), and (iii) constructing estimating functions which allow missing covariate values. The findings from these investigations will be communicated in separate reports.

ACKNOWLEDGEMENTS

This work of D. Y. Lin was supported by the National Institutes of Health, and that of Z. Ying by the National Science Federation and the National Security Agency. The authors are grateful to two referees for their useful comments and to Norman Breslow, Jian Huang, Ian McKeague, Barbara McKnight, Ross Prentice and Jon Wellner for helpful discussions.

References

AALEN, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Lecture Notes in Statistics*, **2**, Ed. N. Klonecki, A. Kosek and J. Rosiński, pp. 1–25. New York: Springer.

AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907–25.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–20.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: John Hopkins Univ.

BRESLOW, N. E. (1972). Discussion of paper of D. R. Cox. *J. R. Statist. Soc.* B **34**, 216–7.

BRESLOW, N. E. & DAY, N. E. (1980). *Statistical Models in Cancer Research*, **1**, *The Design and Analysis of Case-Control Studies*. Lyon: IARC.

BRESLOW, N. E. & DAY, N. E. (1987). *Statistical Methods in Cancer Research*, **2**, *The Design and Analysis of Cohort Studies*. Lyon: IARC.

BUCKLEY, J. D. (1984). Additive and multiplicative models for relative survival rates. *Biometrics* **40**, 51–62.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

COX, D. R. & OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.

HUFFER, F. W. & MCKEAGUE, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Am. Statist. Assoc.* **86**, 114–29.

LAI, T. L. & YING, Z. (1992). Asymptotically efficient estimation in censored and truncated regression models. *Statist. Sinica* **2**, 17–46.

MCKEAGUE, I. W. (1992). Discussion of the paper by P. D. Sasieni. In *Survival Analysis: State of the Art*, Ed. J. P. Klein and P. K. Goel, pp. 263–5. Dordrecht: Kluwer Academic Publishers.

PIERCE, D. A. & PRESTON, D. L. (1984). Hazard function modelling for dose-response analysis of cancer incidence in A-bomb survivor data. In *Atomic Bomb Survivor Data: Utilization and Analysis*, Ed. R. L. Prentice and D. J. Thompson, pp. 51–66. Philadelphia: SIAM.

POCOCK, S. J., GORE, S. M. & KERR, G. R. (1982). Long term survival analysis: the curability of breast cancer. *Statist. Med.* **1**, 93–104.

THOMAS, D. C. (1986). Use of auxiliary information in fitting nonproportional hazards models. In *Modern Statistical Methods in Chronic Disease Epidemiology*, Ed. S. H. Moolgavkar and R. L. Prentice, pp. 197–210. New York: Wiley.

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

[*Received April* 1992. *Revised June* 1993]