#29

Contemporary Mathematics
Volume **80**, 1988

CENSORING, MARTINGALES, AND THE COX MODEL

Ya'acov Ritov and Jon A. Wellner

ABSTRACT. A brief survey of regression models for survival data is given. We then introduce and study two basic operators arising frequently in survival analysis, the (logarithmic) derivative operators (at a distribution function F) corresponding to the mappings from density function $f = F'$ to hazard function $\lambda = f/(1-F)$ and from hazard function λ back to density function f. These operators, which we call R and L respectively, are bounded operators on $L_2(F)$ with $R^T = L$, $L^T = R$, and $L^{-1} = R$, $R^{-1} = L$ for mean 0 functions in $L_2(F)$; hence R and L are both unitary operators. The operators R and L play a fundamental role as links between counting process martingales and the Doob (conditional expectation) martingales which arise via censoring. In the last two sections, we use the properties of the R and L operators to give simple information bound calculations for both the Cox proportional hazards model and for the accelerated life model with censoring.

1. Introduction: Regression models for survival data

The classical Cox (1972) regression model for survival data is

$$\lambda(t\,|\,z) = e^{\theta^T z}\lambda(t) \tag{1.1}$$

where λ is an unknown baseline hazard function, $\lambda = g/(1-G)$ for some absolutely continuous df G with density g, $Z \in \mathbf{R}^k$ is a vector of covariates with distribution H, $\theta \in \mathbf{R}^k$ is the vector of unknown regression parameters, and $\lambda(t\,|\,z)$ denotes the conditional hazard function of survival at t given $Z = z$. Because the model (1.1) is formulated in terms of hazard functions, it lends itself to the use of martingale methods in the study of estimators and calculation of information bounds, and censoring is easily incorporated. See Andersen and Gill (1982) for a study of the Cox partial likelihood estimators via martingales, and see section 4 below for a corresponding martingale derivation of information bounds for estimation of

θ in the model (1.1) which is somewhat easier than the treatment by Begun, Hall, Huang, and Wellner (1983) and the earlier calculations by Efron (1977).

Many variants of the Cox model (1.1) and other regression models for survival data have been proposed during the past few years. We briefly review some of these variations, beginning with other models for the conditional hazard function $\lambda(t \mid z)$.

Hazard function models

Apparently the first attempt to estimate the conditional cumulative hazard function $\Lambda(t \mid z)$ based on censored data and

(1.2) $\quad \lambda(t \mid z)$ arbitrary ,

without any assumptions about its structure, was that of Beran (1981). He used kernel smoothing in the covariate space, and established consistency of the resulting nonparametric estimators. Dabrowska (1987a, 1987b) has studied rates of convergence and asymptotic normality of Beran's estimator.

When $Z = (Z_1, Z_2) \in \mathbf{R}^m \times \mathbf{R}^l$, there are many useful variants of (1.1). A model that lies between (1.1) and (1.2) is

(1.3) $$\lambda(t \mid z) = \exp(\theta_1^T z_1)\, \lambda(t \mid z_2) ;$$

where θ_1 is unknown and $\lambda(t \mid z_2)$ is arbitrary; this is sometimes called the *stratified Cox model* if z_2 takes on only finitely many values. Keeping the proportional hazards form of the model (1.1), but allowing the form of the regression to be an arbitrary (smooth) function, yields the model

(1.4) $$\lambda(t \mid z) = e^{r(z)} \lambda(t)$$

where $r : \mathbf{R}^k \to \mathbf{R}^1$. This form of the model has been studied by O'Sullivan (1986). If k is large, it is often useful to make further assumptions about r; for example $r(z) = \theta_1^T z_1 + r_2(z_2)$ leads to a "partly linear" form of the Cox model analogous to the partly linear generalizations of the usual regression model studied by Engle, Granger, Rice, and Weiss (1986), Green (1985), Green and Yandell (1985), and Heckman (1986):

(1.5) $$\lambda(t \mid z) = \exp(\theta_1^T z_1 + r_2(z_2))\, \lambda(t) .$$

If r is assumed to be additive, $r(z) = \sum_{i=1}^{k} r_i(z_i)$ where

$r_i : \mathbf{R}^1 \to \mathbf{R}^1$ are arbitrary, the resulting models are analogous to the additive forms of the usual regression models; see e.g. Hastie and Tibshirani (1987).

A different generalization of the model (1.1) was proposed by Aalen (1978), (1980): in Aalen's model

$$\lambda(t \mid z) = \sum_{i=1}^{k} z_i \lambda_i(t) \tag{1.6}$$

where the λ_i are unknown functions. In fact, Aalen allows the Z_i's to be predictable processes $Z_i(t)$. McKeague (1985) has studied estimation of the λ_i's via sieves. Another variation of this type of model (which avoids the non-negativity constraints implicit in (1.6)) is

$$\lambda(t \mid z) = \exp(\sum_{i=1}^{k} z_i \alpha_i(t)) \tag{1.7}$$

where the functions α_i are unknown. This model was introduced and studied by Zucker (1986) and Zucker and Karr (1987).

Generalizations of (1.1) which allow relative risk functions other than $\exp(\theta^T z)$ have been proposed and studied by Thomas (1981), Breslow and Storer (1985), Prentice and Self (1983), and Moolgavkar and Venzon (1987). Thomas (1981) replaces $\exp(\theta^T z)$ by

$$\exp[\beta\theta^T z + (1-\beta)\log(1+\theta^T z)^+]$$

with $0 \le \beta \le 1$, while Breslow and Storer (1985) replace $\exp(\theta^T z)$ by

$$\begin{cases} \exp\{\dfrac{(1+\theta^T z)^\beta - 1}{\beta}\}, & \beta \ne 0 \\ (1+\theta^T z)^+, & \beta = 0. \end{cases}$$

Prentice and Self (1983) also allow time dependent covariates.

Transformation models

The Cox model (1.1) can also be written as

$$s(T) = -\theta^T Z + \varepsilon \tag{1.8}$$

where $s \equiv \log \Lambda$ and θ are unknown and $e^{\varepsilon} \sim$ exponential(1), so $\varepsilon \sim$ extreme value: $P(\varepsilon \geq t) = P(e^{\varepsilon} \geq e^t) = \exp(-e^t)$. (This is easily obtained from (1.1) by writing it in terms of cumulative hazards as $\Lambda(t \mid z) = \exp(\theta^T z)\Lambda(t)$, evaluating this at T and Z, and noting that $\Lambda(T \mid Z) \sim$ exponential(1).) If ε has some (known) distribution, perhaps not extreme value, then (1.8) is called a *transformation model.* This type of model arises if the underlying model is a Cox model as in (1.1), but some covariate Z_i is not observed; see Clayton and Cuzick (1986). The theory of efficient estimation for this type of model is difficult, and has not yet incorporated censoring. For the best results to date, see Bickel (1986).

Accelerated life models

Another alternative to the Cox regression model (1.1) is the accelerated life model

$$\tilde{g}(t \mid z) = e^{-\theta^T z}\,\tilde{g}(e^{-\theta^T z} t) \tag{1.9}$$

where $\tilde{g}$ is an unknown density function, $\theta \in \mathbf{R}^k$ is unknown, and $\tilde{g}(t \mid z)$ is the conditional density of T given $Z = z$. Transformation to $Y = \log T$ gives

$$g(y \mid z) = g(y - \theta^T z) \tag{1.10}$$

with $g(y) = e^y \tilde{g}(e^y)$ unknown, or

$$Y = \theta^T Z + \varepsilon, \qquad \text{where } \varepsilon \sim g. \tag{1.11}$$

Thus the accelerated life model can be transformed to an ordinary linear regression model. Note that the model (1.9) coincides with the Cox model (1.1) if $\tilde{g}$ is a Weibull density. See e.g. Cox and Oakes (1984).

The problem is that censoring creates more difficulties here since there is no immediate martingale structure in the basic model (1.11). These difficulties have hindered wide use of this model for survival analysis. In section 5 below we show how censoring induces a martingale structure in the problem which can be exploited to give a relatively clean and straightforward derivation of information bounds for estimation of θ in the model (1.11) based on censored observations. Our derivation also yields the natural family of estimating equations developed by Tsiatis (1986) from consideration of rank tests.

Estimates of θ in this model have been previously proposed and studied by Buckley and James (1979), James and Smith (1984), and by Tsiatis (1986) who also makes nice use of martingale methods. Their estimates are, in general, inefficient. Efficient estimates have been proposed and studied by Ritov (1984).

Discussion and Summary

Many of the alternative models to the Cox model (1.1) reviewed above involve additional parameters or have been difficult to study with censoring. Our purpose in the remainder of this paper is to introduce two basic operators which occur repeatedly in the survival analysis and reliability literature, but have not been isolated and systematically studied. In section 2 we introduce the two basic operators R and L and explore their properties. (These operators have been independently discovered by Efron and Johnstone (1987); see remark 2.4.) In section 3 we give connections between these operators and martingales which occur naturally in the analysis of censored data. Systematic use of the results of sections 2 and 3 yields relatively straightforward semiparametric information calculations for the parametric components of both the Cox model (1.1) and the accelerated life model (1.11) with random censoring. We present these calculations in sections 4 and 5. Efficiency calculations for the Kaplan - Meier estimator alternative to those of Wellner (1982) are also easily carried out with the help of the results for the R and L operators given in sections 2 and 3; these will be given in Bickel, Klaassen, Ritov, and Wellner (1988).

These methods will also be very useful in treating many of the other models reviewed above. In fact, information calculations for the parametric components of models (1.3) and (1.5) have already been carried out by Peter Sasieni at the University of Washington, and will appear in Sasieni (1988), along with constructions of efficient estimators.

2. The R and L operators and their properties

We first explore the relationships between a collection of density functions $\{f_\theta : \theta \in \Theta \subset \mathbf{R}^1\}$ on $\mathbf{R}^1$ and the corresponding hazard functions $\{\lambda_\theta : \theta \in \Theta \subset \mathbf{R}^1\}$. It is well - known that

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \equiv (Hf)(t) \tag{2.1}$$

where $1 - F(t) \equiv \int_t^\infty f(s)\,ds = P[X \geq t]$; we regard H as a mapping from the set of density functions to the set of hazard functions. Similarly, it is well-known that

$$f(t) = \lambda(t)\exp(-\int_0^t \lambda(s)\,ds) \equiv (D\lambda)(t); \tag{2.2}$$

here we regard D as a mapping from the set of hazard functions to the set of density functions on $\mathbf{R}^1$.

It is really the derivative mappings corresponding to H and D which concern us here. First consider (2.1) for a smooth family $\{f_\theta : \theta \in \Theta\}$. Taking logarithms across (2.1), differentiating with respect to θ, and setting $a(t) \equiv \frac{\partial}{\partial\theta}\log f_\theta(t)$, yields

$$\frac{\partial}{\partial\theta}\log\lambda_\theta(t) = a(t) - \frac{\int_t^\infty a\,dF}{1 - F(t)} \equiv (Ra)(t). \tag{2.3}$$

Note that

$$\begin{aligned} Ra(t) &= -E\{a(X) - a(t) \mid X > t\} \\ &= -\text{residual life of } a(X) \text{ at } t. \end{aligned} \tag{2.4}$$

Since $e(t) \equiv E(X - t \mid X > t)$ is the mean residual life of X at t, we sometimes write $Ra(t) = -e_a(t)$.

Now consider (2.2) for a smooth family $\{\lambda_\theta : \theta \in \Theta \subset \mathbf{R}^1\}$. Taking logarithms across (2.2), differentiating with respect to θ, and setting $b(t) \equiv \frac{\partial}{\partial\theta}\log\lambda_\theta(t)$ yields, with

$$\Lambda(t) \equiv \int_{-\infty}^t \lambda(s)\,ds \equiv \int_{-\infty}^t (1 - F(s))^{-1}\,dF(s),$$

$$\frac{\partial}{\partial\theta}\log f_\theta(t) = b(t) - \int_{-\infty}^t b(s)\lambda(s)\,ds \tag{2.5}$$

$$= b(t) - \int_{-\infty}^{t} b \, d\Lambda$$

$$= b(t) - \int_{-\infty}^{\infty} 1_{[t \geq s]} \, b(s) \, d\Lambda(s)$$

$$\equiv (Lb)(t) .$$

Note that L is a "martingale operator": if $X \sim F$ and

(2.6) $$\boldsymbol{M}(t) = 1_{[X \leq t]} - \int_{-\infty}^{t} 1_{[X \geq s]} \, d\Lambda(s)$$

is the counting process martingale, then

(2.7) $$Lb(X) = \int_{-\infty}^{\infty} b \, d\boldsymbol{M} .$$

Both $R \equiv R(F)$ and $L \equiv L(F)$ are bounded operators from $L_2(F)$ to $L_2(F)$ as will be shown below; throughout most of the following we will suppose that the df F is continuous and suppress the dependence of R and L on F. Even more interesting is that both the adjoint $L^T = R$ and the inverse $L^{-1} = R$ on $L_2^0(F) \equiv \{ a \in L_2(F) : E\, a(X) = \int a \, dF = 0 \}$: by straightforward calculation

(2.8) $$L \circ Ra = a - E\, a(X) \qquad \text{for } a \in L_2(F)$$

and

(2.9) $$R \circ La = a \qquad \text{for } a \in L_2(F) .$$

This is easily understood in the case $E\, a(X) = 0$ by writing

$$f_\theta(t) = (D \circ H)(f_\theta(t)) ,$$

and then taking logarithmic derivatives as in (2.3) and (2.5) to obtain

$$a(t) = L \circ Ra(t) .$$

Similarly, writing

$$\lambda_\theta(t) = (H \circ D) \lambda_\theta(t)$$

and taking (logarithmic) derivatives yields (2.9). To see that the adjoint of L is R and vice versa, let $<\cdot, \cdot>$ denote the inner product in $L_2(F)$, and let $a, b \in L_2(F)$. Then Fubini's theorem yields

$$<La, b> = \int_{-\infty}^{\infty} \{ a(t) - \int_{-\infty}^{t} a\, d\Lambda \} b(t)\, dF(t)$$

$$= \int_{-\infty}^{\infty} a(t) \{ b(t) - \frac{\int_t^{\infty} b\, dF}{1 - F(t)} \} dF(t)$$

$$= <a, Rb>, \tag{2.10}$$

and hence $L^T = R$. The conclusion of these calculations is that the mappings R and L are isometries of $L_2^0(F)$. Let $\|\cdot\|$ denote the norm in $L_2(F)$: $\|a\|^2 = <a, a>$. Then for $a \in L_2^0(F)$ we have

$$\begin{aligned} \|La\|^2 &= <La, La> \\ &= <a, L^T La> && \text{by definition of } L^T \\ &= <a, RLa> && \text{by } L^T = R \\ &= <a, a> = \|a\|^2 && \text{by the identity (2.8).} \end{aligned} \tag{2.11}$$

A similar calculation holds for $\|Ra\|^2$ for $a \in L_2^0(F)$, but more generally, by (2.7) it follows that

$$\begin{aligned} Var[a(X)] &= <a, a - Ea> \\ &= <a, LRa> = <a, R^T Ra> \\ &= <Ra, Ra> = \|Ra\|^2 . \\ &= E[a(X) - A(X)]^2 \end{aligned} \tag{2.12}$$

where

$$A(t) \equiv E[a(X) | X > t]. \tag{2.13}$$

We summarize these results in the following proposition:

Proposition 2.1. Suppose that the df F is continuous. The operators R and L mapping from $L_2(F)$ to itself defined by (2.3) and (2.5) respectively satisfy:

(i) R and L are bounded with $\|R\| = \|L\| = 1$.

(ii) $L \circ Ra = a - E\,a(X)$ and $R \circ La = a$; thus $R^{-1} = L$ on

$$\mathbf{H}_0 \equiv L_2^0(F) = \{a \in L_2(F) : E\,a = 0\}.$$

(iii) $L = R^T$ and $R = L^T$; hence L and R are isometries of $\mathbf{H}_0$ (or unitary transformations): $\|La\| = \|a\| = \|Ra\|$ for all $a \in \mathbf{H}_0$ and $L^T L = R^T R =$ identity on $\mathbf{H}_0$.

(iv) $Var[a(X)] = E[Ra(X)]^2 = E[e_a(X)]^2$ or, with A defined in (2.13),

$$Var[a(X)] = E[a(X) - A(X)]^2.$$

Proof. It remains only to prove (i). Since $L^T = R$, by the theory of adjoints it suffices to prove that one of R and L is bounded. We will give two different proofs. The first shows that L is bounded by a martingale argument, while the second uses Hardy's inequality to show that R is bounded.

First proof of (i): For $a \in L_2(F)$ set

(a) $$\mathbb{Z}(t) \equiv \int_{-\infty}^{t} a\, d\boldsymbol{M}$$

where $\boldsymbol{M}$ is the counting process martingale of (2.6) with predictable variation process $<\boldsymbol{M}>(t) = \int_{-\infty}^{t} 1_{[X \ge s]}\, d\Lambda(s)$. Then $\mathbb{Z}$ is a square integrable martingale with

(b) $$\mathbb{Z}(t) \to_{a.s.} La(X) \quad \text{as } t \to \infty,$$

and with predictable variation process

$$<\mathbb{Z}>(t) = \int_{-\infty}^{t} a^2\, d<\boldsymbol{M}>.$$

Thus

$$\begin{aligned} E[\mathbb{Z}^2(t)] &= E[<\mathbb{Z}>(t)] \\ &= E[\int_{-\infty}^{t} a^2\, d<\boldsymbol{M}>] \\ &= E[\int_{-\infty}^{t} a^2(s)\, 1_{[X \ge s]}\, d\Lambda(s)] \\ \text{(c)} \qquad &= \int_{-\infty}^{t} a^2(s)\, dF(s). \end{aligned}$$

Hence by Fatou's lemma and (c),

(d) $$E\{[La(X)]^2\} = E\{\lim_{t \to \infty} \mathbb{Z}^2(t)\}$$

$$\leq \liminf_{t \to \infty} \int_{-\infty}^{t} a^2 \, dF$$

$$= E[a^2(X)] < \infty .$$

If follows that L is bounded (as an operator from $L_2(F)$ to $L_2(F)$) and $\|L\| \leq 1$.

The second proof of (i) uses Hardy's inequality. This proof was pointed out to us by Peter Bickel. Let $T : L_2(0,1) \to L_2(0,1)$ be defined by

$$\text{(e)} \qquad Th(x) = \frac{1}{x} \int_0^x h(y)\, dy \qquad \text{for} \quad h \in L_2(0,1) .$$

Then T is bounded with $\|T\| = 2$: i.e.

$$\text{(f)} \qquad \int_0^1 \{ \frac{1}{x} \int_0^x h(y)\, dy \}^2 \, dx \leq 4 \int_0^1 h^2(y)\, dy .$$

This is Hardy's inequality; see Hardy, Littlewood, and Polya (1952, second ed.) page 240, Dunford and Schwartz (1958) page 582, or Rudin (1973) page 107.

Boundedness of $Ra = a - \int_{\cdot}^{\infty} a \, dF/(1 - F)$ follows from boundedness of T by the probability integral transformation: since the first term of R is clearly bounded, it suffices to show that

$$\text{(g)} \qquad \int_{-\infty}^{\infty} \{ \frac{1}{1 - F(t)} \int_t^{\infty} a \, dF \}^2 \, dF(t) \leq 4 \int_{-\infty}^{\infty} a^2 \, dF .$$

But by letting $u = 1 - F(t) = \overline{F}(t)$, the left side of (g) equals

$$\int_0^1 \{ \frac{1}{u} \int_0^u a(F^{-1}(1-s))\, ds \}^2 du$$

$$\leq 4 \int_0^1 [a(F^{-1}(1-s))]^2 \, ds \qquad \text{by (f)}$$

$$= 4 \int_{-\infty}^{\infty} a^2 \, dF .$$

Thus (g) holds and R is bounded. □

We conclude this section with some remarks on proposition 2.1.

Remark 2.1. The equality $\|La\| = \|a\|$ in (iii) follows from the martingale representation (7) of L and (i).

Remark 2.2. The variance identity (iv) generalizes the formula

$$(2.14) \qquad Var[X] = E[e^2(X)]$$

for $X \sim F$ continuous noted by Pyke (1965), and extended to the case of discontinuous F by Hall and Wellner (1981): they show that

$$(2.15) \qquad Var[X] = E[e(X)\, e(X-)]$$

always holds. These formulas were extended to any $a \in L_2(F)$ and arbitrary F by Shorack and Wellner (1986), page 283: from (2.7) and (2.8) it follows that

$$(2.16) \qquad a(X) - E\,a(X) = \int_{-\infty}^{\infty} Ra\, d\mathbf{M} ,$$

and hence a martingale calculation yields

$$(2.17) \qquad \begin{aligned} Var[a(X)] &= E[(Ra)^2(X)(1 - \Delta\Lambda(X))] \\ &= E[e_a^2(X)\,(1 - \Delta\Lambda(X))] . \end{aligned}$$

(Note that $1 - \Delta\Lambda(x) = (1 - F(x))/(1 - F(x-))$.)

Remark 2.3. Note that our first proof of (i) gives an indirect proof (via adjoints) of Hardy's inequality. Csorgo, Csorgo, and Horvath (1986) page 40 also use Hardy's inequality in connection with estimation of mean residual life (as suggested by David Mason).

Remark 2.4. These and other properties of R and L have been independently discovered by Efron and Johnstone (1987). (Their A is our R and their B is our L .) They point out the following consequence of proposition 2.1(v). For a smooth family of functions $\{h_\theta : \theta \in \Theta \subset \mathbf{R}^1\}$, write

$$\dot{h}_\theta(t) \equiv \frac{\partial}{\partial\theta} h_\theta(t) .$$

Suppose that $\{f_\theta : \theta \in \Theta\}$ is a smooth family of functions with $\dot{f}_\theta / f_\theta \in L_2(F_\theta) \equiv L_2(F)$. Then it follows immediately from (3) that

$$\frac{\dot{\lambda}_\theta}{\lambda_\theta}(t) = \frac{\partial}{\partial\theta} \log\lambda_\theta(t) = R(\frac{\dot{f}_\theta}{f_\theta})(t) ,$$

and hence (v) yields an identity for the Fisher information I_θ for θ :

$$(2.18) \qquad I_\theta \equiv E_\theta[(\frac{\dot{f}_\theta}{f_\theta})^2] = E_\theta[(\frac{\dot{\lambda}_\theta}{\lambda_\theta})^2] .$$

3. Martingale connections

The R and L operators play an important role as links between the martingales which arise via censoring (Doob's martingales) and the counting process martingale M of (2.6).

Let $X \sim F$ be defined on a probability space $(\Omega, \mathbf{F}, P)$, and consider the filtration $\{ \mathbf{F}_t \}_{t \geq 0}$ defined by

$$\mathbf{F}_t \equiv \sigma\{ 1_{[X \leq s]} : s \leq t \} = \sigma\{ X \wedge t , 1_{[X \leq t]} \} \tag{3.1}$$

for $t \geq 0$. For a fixed function $a \in L_2(F)$ and $t \geq 0$ set

$$\mathbf{Y}(t) \equiv E[a(X) | \mathbf{F}_t] . \tag{3.2}$$

If $a \equiv (\partial/\partial\theta)\log q_\theta$ is a score function for estimation of a parameter θ in a model $\mathbf{Q} \equiv \{q_\theta : \theta \in \Theta \subset \mathbf{R}^1 \}$ based on observation of X, then $\mathbf{Y}(t)$ is the score function for estimation of θ in the induced model $\mathbf{P} \equiv \{p_\theta : \theta \in \Theta \}$ based on the censored observation $T \equiv X \wedge t$.

It is well-known that $\{ \mathbf{Y}(t), \mathbf{F}_t \}_{t \geq 0}$ is a (uniformly - integrable) martingale for any increasing filtration $\mathbf{F}_t$ and integrable function a. This is sometimes called *Doob's martingale*; see e.g. Karlin and Taylor (1975) page 246 for the discrete time version of this, and Elliott (1982) page 36 for the continuous time result. Conversely, every uniformly integrable martingale can be written as the conditional expectation of some integrable function a as in (3.2); see e.g. Elliott (1982) page 36, or Liptser and Shiryayev (1978), theorem 2.7, page 45. For our present particular filtration $\{\mathbf{F}_t\}$ it is easily shown that

$$\begin{aligned} \mathbf{Y}(t) &= 1_{[X \leq t]} a(X) + 1_{[X > t]} \frac{E\, a(X) 1_{[X > t]}}{1 - F(t)} \\ &= 1_{[X \leq t]} a(X) + 1_{[X > t]} A(t) \end{aligned} \tag{3.3}$$

with A as in (2.13).

Now we can relate the martingale $\mathbf{Y}$ to the counting process martingale

$$\begin{aligned} M(t) &\equiv N(t) - \mathbb{A}(t) \\ &= 1_{[X \leq t]} - \int_{-\infty}^{t} 1_{[X \geq s]}\, d\Lambda(s) \end{aligned} \tag{3.4}$$

defined in (2.6). The following proposition is apparently due to Chou and Meyer (1974) and (1975): see theorem 2 and formulas (9) and (10) of Chou and Meyer (1974), page 1563; and see proposition 2 and formula (13) of Chou and Meyer (1975), pages 231 - 232. (The (q, M, H, h) of Chou and Meyer (1975) correspond to our our $(\mathbb{M}, \mathbb{Y}, a, Ra)$. Their result is somewhat more general in that they only require $a \in L_1(F)$.) For more on related martingale representation theorems see Liptser and Shiryayev (1978), chapter 19.

Proposition 3.1. Suppose that $a \in L_2(F)$ and $E\,a(X) = 0$. Then the martingale $\mathbb{Y}(t)$ in (3.2) and (3.3) is related to the counting process martingale $\mathbb{M}$ in (3.4) by

$$(3.5) \qquad \mathbb{Y}(t) = \int_{-\infty}^{t} Ra(s)\, d\mathbb{M}(s)$$

where, as in (2.3),

$$(3.6) \qquad Ra(t) \equiv a(t) - \frac{\int_t^\infty a\, dF}{1 - F(t)} = a(t) - A(t).$$

Proof. First write $\mathbb{M} = \mathbb{N} - \mathbb{A}$ and note that since $Ra = a - A$,

$$(a) \qquad \int_{-\infty}^{t} Ra\, d\mathbb{N} = a(X)1_{[X \le t]} - A(X)1_{[X \le t]}.$$

But then

$$(b) \qquad \int_{-\infty}^{t} Ra\, d\mathbb{A} = \int_{-\infty}^{t \wedge X} Ra\, d\Lambda$$
$$= -A(X)1_{[X \le t]} - A(t)1_{[X > t]}$$

since, by a calculation using Fubini's theorem and $\int a\, dF = 0$ twice,

$$\int_{-\infty}^{u} Ra\, d\Lambda = \int_{-\infty}^{u} a\, d\Lambda - \int_{-\infty}^{t} \{ \frac{1}{1 - F(s)} \int_s^\infty a\, dF \}\, d\Lambda(s)$$
$$= -A(u).$$

Subtracting (a) and (b) yields (3.5). $\square$

Remark 3.1. If a in (3.3) and (3.5) is replaced by $a - E\,a(X)$ for an arbtrary $a \in {}_2(F)$, then $R(a - E\,a(X)) = Ra$ and (3.3) together with (3.5) imply that

$$a(X) - E\,a(X) = \boldsymbol{Y}(\infty) = L(Ra)(X)$$

in agreement with (2.8).

Remark 3.2. The predictable variation process of the martingale transform $\boldsymbol{Y}$ is

$$<\boldsymbol{Y}>(t) = \int_{-\infty}^{t} (Ra)^2(s)\, 1_{[X \geq s]}\, d\Lambda(s) \tag{3.7}$$

so that both

$$E[\boldsymbol{Y}^2(t)] = E<\boldsymbol{Y}>(t) = \int_{-\infty}^{t} (Ra)^2\, dF \tag{3.8}$$

from (3.7) and

$$E[\boldsymbol{Y}^2(t)] = \int_{-\infty}^{t} a^2\, dF + A^2(t)(1 - F(t)) \tag{3.9}$$

directly from (3.3).

Now let $X_1, X_2, \cdots$ be iid F and, for $t \in R^1$, set

$$\boldsymbol{F}_n(t) = n^{-1} \sum_{i=1}^{n} 1_{[X_i \leq t]}\,, \tag{3.10}$$

$$\boldsymbol{X}_n(t) \equiv \sqrt{n}\,[\boldsymbol{F}_n(t) - F(t)]\,, \tag{3.11}$$

and

$$\boldsymbol{M}_n(t) \equiv \sqrt{n}\,[\boldsymbol{F}_n(t) - \int_{-\infty}^{t} (1 - \boldsymbol{F}_n(s-))\, d\Lambda(s)]\,; \tag{3.12}$$

$\boldsymbol{F}_n$ is the empirical df, $\boldsymbol{X}_n$ is the empirical process, and $\boldsymbol{M}_n$ is the normalized counting process martingale corresponding to the counting process $n\boldsymbol{F}_n$. While both $\boldsymbol{X}_n$ and $\boldsymbol{M}_n$ have mean zero, $\boldsymbol{X}_n$ records unconditional deviations of $\boldsymbol{F}_n$ from the true F, whereas $\boldsymbol{M}_n$ records conditional deviations of $\boldsymbol{F}_n$ from the true df F.

It is useful to relate these two sets of deviations. From (2.16), or from (3.5) with $t = \infty$, it follows that

$$a(X_i) - Ea(X) = \int_{-\infty}^{\infty} Ra\, d\boldsymbol{M}_i \tag{3.13}$$

with

$$M_i(t) \equiv 1_{[X_i \le t]} - \int_{-\infty}^{t} 1_{[X_i \ge s]} \, d\Lambda(s) .$$

Summing (3.13) on i and dividing by $n^{1/2}$ yields

$$\int a \, d\mathbf{X}_n = \int Ra \, d\mathbf{M}_n , \tag{3.14}$$

or equivalently, since $R \circ L = I$, by taking $a = Lb$,

$$\int Lb \, d\mathbf{X}_n = \int b \, d\mathbf{M}_n . \tag{3.15}$$

Since $\mathbf{X}_n =_d U_n(F) \Rightarrow B^0(F)$ as $n \to \infty$ where U_n is the empirical df of $\xi_1, \cdots, \xi_n$ iid Uniform(0,1) and B^0 is a Brownian bridge process, and $M_n \Rightarrow B(F)$ where B is a standard Brownian motion, as $n \to \infty$ (3.14) and (3.15) become

$$\int a \, d\mathbf{B}^0(F) =_d \int Ra \, dB(F) \tag{3.16}$$

and

$$\int Lb \, d\mathbf{B}^0(F) =_d \int b \, d\mathbf{B}(F) . \tag{3.17}$$

Thus R and L give a way of relating integrals with respect to the empirical process $\mathbf{X}_n$ (unconditional deviations) to integrals with respect to the martingale process M_n (conditional deviations). See Shorack and Wellner (1986) chapter 6 for these and other connections between $\mathbf{X}_n$ and M_n and between the limit processes $B^0(F)$ and $B(F)$.

4. Information Calculations for the Cox model

Because the Cox model is formulated in terms of hazards, martingales appear naturally in the basic model even without any censoring, and then the addition of censoring creates little or no real additional complication. Here we will give a complete derivation of the information bounds for estimation of the regression parameter θ in the Cox model (1.1) without censoring. These calculations show how the L and R operators arise naturally in many situations of this kind. While the results are exactly the same as those presented in Begun, Hall, Huang, and Wellner (1983), the present calculations are both simpler and more complete.

We assume, as in (1.1), that we observe (Z, T) with conditional hazard function

(4.1) $\lambda(t \mid z) = r(\theta z)\lambda(t) \equiv \lambda_r(t)$

where $r(z) \equiv e^z$, and that $Z \sim H$ is, for simplicity, real-valued. (We also suppose, for simplicity, that H is known; if H is unknown, it can be shown that the information for θ is unchanged.) Thus the joint density of the observations is, with $r \equiv r(\theta z)$,

(4.2) $f(z, t) = r\, g(t) \overline{G}^{r-1}(t) h(z)$.

Straightforward calculation yields the score for θ:

(4.3) $\dot{l}_1(z, t) = z[1 - \Lambda_r(t)]$

where

(4.4) $\Lambda_r(t) \equiv r(\theta z)\Lambda(t) = r(\theta z)\int_0^t \frac{dG}{1-G}$.

Similarly, letting $\{g_\eta : \eta \in \mathbf{R}^1\}$ be a regular parametric family and $a \equiv \frac{\partial}{\partial\eta}\log g_\eta$, the score (operator) for g is

(4.5) $\dot{l}_2 a(z, t) = a(t) + (r-1)\frac{\int_t^\infty a\, dG}{1 - G(t)}$.

Now we have two sets of L and R operators: L and R corresponding to the survival function $1 - G$ with hazard rate λ; and, conditional on $Z = z$, L_r and R_r corresponding to $(1-G)^r$ with hazard rate $\lambda_r \equiv r\lambda$. Thus

$$L_r a(t) = a(t) - \int_0^t a\, d\Lambda_r = a(t) - r\int_0^t a\, d\Lambda$$

and

$$R_r a(t) = a(t) - \frac{\int_t^\infty a\, dF(\cdot \mid z)}{1 - F(t \mid z)}$$

$$= -E\{a(T) - a(t) \mid Z = z, T > t\}.$$

The scores $\dot{l}_1$ and $\dot{l}_2$ are easily expressed in terms of L_r and R: we have

(4.6) $\dot{\mathrm{l}}_1(z, t) = z\,(L_r 1)(t)$

while

(4.7) $\dot{\mathrm{l}}_2\, a(z, t) = (L_r\, R\, a)(z,t)\,.$

The formula (4.7) follows since the right side of (4.5) equals

$$Ra(t) + r\,\frac{\int_t^\infty a\, dG}{1 - G(t)} = Ra(t) - r\int_0^t Ra\, d\Lambda$$

by (c) of the proof of proposition 3.1

$$= Ra(t) - \int_0^t Ra\, d\Lambda_r$$

$$= L_r\, Ra(t)\,.$$

Thus by proposition 3.1,

(4.8) $\dot{\mathrm{l}}_1(Z, T) = Z\int_0^\infty dM_r(s)$

and

(4.9) $\dot{\mathrm{l}}_2\, a(Z, T) = \int_0^\infty Ra(s)\, dM_r(s)$

where

(4.10) $M_r(t) \equiv 1_{[T \le t]} - \int_0^t 1_{[T \ge s]}\, d\Lambda_r(s)\,.$

To calculate the efficient score function l_1^* for θ, we want to find a function a^* with $\int a^*\, dG = 0$ so that

$$\mathrm{l}_1^* \equiv \dot{\mathrm{l}}_1 - \dot{\mathrm{l}}_2\, a^* \perp \dot{\mathrm{l}}_2\, a \qquad \text{in } L_2(P)$$

for all functions $a \in L_2^0(P)$; i.e.

(4.11) $E\{[\dot{\mathrm{l}}_1 - \dot{\mathrm{l}}_2 a^*]\,\dot{\mathrm{l}}_2 a\} = 0 \qquad \text{for all } a \in L_2^0(P)\,.$

This is just as in Begun, Hall, Huang, and Wellner (1983), except that here we are working in $L_2(P)$ rather than $L_2(\mu)$ and have replaced A by

$\dot{\mathbf{l}}_2$, β^* by a^*, and β by a.

By conditioning on Z, the expectation in (4.11) is easily calculated as the expectation of the predictable covariation process of the martingale transforms in (4.8) and (4.10): thus the left side of (4.11) equals

$$
\begin{aligned}
E\,E\{ [\dot{\mathbf{l}}_1 - \dot{\mathbf{l}}_2 a^*]\dot{\mathbf{l}}_2 a \mid Z\} \\
&= E\,E\{ \int (Z - Ra^*) Ra\, 1_{[T \geq s]}\, r(\theta Z)\, d\Lambda(s) \mid Z \} \\
&= \int \{ E[Z\, r(\theta Z)\, 1_{[T \geq s]}] \\
&\qquad - E[r(\theta Z)\, 1_{[T \geq s]}]\, Ra^*(s) \}\, Ra(s)\, d\Lambda(s) \\
&\equiv \int \{ S_1(s) - S_0(s) Ra^*(s)\}\, Ra(s)\, d\Lambda(s) \qquad (4.12)
\end{aligned}
$$

where

$$
S_i(t) \equiv E\{Z^i\, r(\theta Z)\, 1_{[T \geq t]}\}, \qquad \text{for } i = 0, 1. \qquad (4.13)
$$

From (4.12) it is easy to make the right choice of a^*: set

$$
a^* \equiv L\left(\frac{S_1}{S_0}\right). \qquad (4.14)
$$

Since $R \circ L$ = identity by proposition 2.1, it follows that

$$
Ra^* = \frac{S_1}{S_0},
$$

and hence the integrand of (4.12) is zero identically, and (4.11) holds. Thus the efficient score function for estimation of θ is

$$
\begin{aligned}
\mathbf{l}_1^*(Z, T) &= \dot{\mathbf{l}}_1(Z, T) - \dot{\mathbf{l}}_2 a^*(Z, T) \\
&= \int_0^\infty \left[Z - \frac{S_1(t)}{S_0(t)} \right] d\boldsymbol{M}_r(t) \\
&= \int_0^\infty [Z - E(Z \mid T = t)]\, d\boldsymbol{M}_r(t) \qquad (4.15)
\end{aligned}
$$

since

$$
\frac{S_1}{S_0}(t) = E(Z \mid T = t)
$$

by straightforward calculations. Hence the information for θ is, by an easy martingale calculation

$$I(\theta) = E[\dot{l}_1^*(Z,T)^2]$$

$$= E\,E\{\int_0^\infty [Z - E(Z\,|\,T=t)]^2\, 1_{[T \geq t]}\, r(\theta Z)\, d\Lambda(t)\,|\,Z\}$$

$$= E\int_0^\infty [Z - E(Z\,|\,T=t)]^2\, r(\theta Z)\, \overline{G}(t)^{r(\theta Z)-1}\, dG(t)$$

(4.16)
$$= E[Z - E(Z\,|\,T)]^2 .$$

These calculations also lead to estimating equations which can be used to obtain a whole class of inefficient estimates of θ in the Cox model as follows: Let $J(z,t)$ be a fixed measurable function satisfying $E[J^2(Z,T)] < \infty$, and define

$$h(Z,T) \equiv \int_0^\infty [J(Z,t) - E(J(Z,t)\,|\,T=t)]\, d\boldsymbol{M}_r(t)$$

(4.17)
$$= \int_0^\infty [J(Z,t) - \frac{S_J(t)}{S_0(t)}]\, d\boldsymbol{M}_r(t)$$

where

$$S_J(t) \equiv E\{J(Z,t)\, r(\theta Z)\, 1_{[T \geq t]}\} .$$

By a calculation as in (4.12), it is easily verified that $h(Z,T) \perp \dot{l}_2 a$ for all $a \in L_2^0(G)$. Then

$$0 = E\,h(Z,T) = E\{J(Z,T)\} - E\{\frac{S_J}{S_0}(T)\} ,$$

so by defining

(4.18)
$$W(\theta, P) \equiv E_P\{J(Z,T) - \frac{S_J(T,\theta,P)}{S_0(T,\theta,P)}\} ,$$

$\sqrt{n}$ – consistent estimates $\hat{\theta}_J$ of θ can be defined as the solutions of

(4.19)
$$W(\hat{\theta}_J, \boldsymbol{P}_n) = 0 .$$

This family of estimators will be studied elsewhere. Note that with $J(z,t) = z$, (4.18) and (4.19) yield the Cox partial likelihood estimator.

5. Information calculations for censored regression models

We now illustrate the usefulness of the results of sections 2 and 3 in treating linear regression models when the observations on the dependent variable are subject to (arbitrary right) censoring. For simplicity, we suppose that the independent variable (covariate) is one - dimensional and we treat only information bounds for estimation of the regression parameter θ.

Let $X = (Y, Z, C) \in \mathbf{R}^3$ where

$$Y = \theta Z + \varepsilon \tag{5.1}$$

with ε independent of (Z, C). We suppose that the distribution function F of the error ε has density f. Thus the model $\mathbf{Q}$ is

$$\mathbf{Q} = \left\{ \begin{array}{l} Q : \dfrac{dQ}{d\mu}(y,z,c) = f(y - \theta z)\, h(z,c) \\ \text{for some } f \text{ and } h \text{ and } \theta \in \mathbf{R}^1 \end{array} \right\} . \tag{5.2}$$

We observe

$$T(\underline{X}) = (Z, Y \wedge C, 1_{[Y \le C]}) \equiv (Z, V, \Delta), \tag{5.3}$$

and denote the induced model by $\mathbf{P} \equiv \mathbf{Q}T^{-1}$.

This model arises in survival analysis when Y is the log of failure time, and is sometimes called the "accelerated time model"; see e.g. Kalbfleisch and Prentice (1981) or Lawless (1982). Estimators of θ in this model have been proposed by Miller (1976), Buckley and James (1979), Koul, Susarla and Van Ryzin (1981), and Tsiatis (1986). Information calculations and a construction of efficient estimators were carried out by Ritov (1985).

Here we rederive the information bound calculated by Ritov (1985) using the results of sections 2 and 3, and show how these calculations provide an alternative rationale for the (inefficient) rank estimators proposed by Tsiatis (1986).

We begin by calculating scores for θ and f in the model $\mathbf{Q}$: by straightforward differentiation, the score for θ is

$$\dot{\mathrm{l}}_1(y,z;\mathrm{Q}) = -z\,\frac{f'}{f}(y - \theta z) = z\,\psi(\varepsilon) \tag{5.4}$$

where $\psi(x) \equiv -f'(x)/f(x)$. To calculate the score for f, let $\{f_\eta : \eta \in \mathbf{R}^1\}$ be a regular parametric family and set $a \equiv \frac{\partial}{\partial\eta}\log f_\eta$. Then the score (operator) for f is

(5.5) $$\dot{\mathrm{l}}_2\, a(y,z;\mathrm{Q}) = a(y-\theta z) = a(\varepsilon)\,.$$

To calculate scores in the induced model $\mathbf{P} = \mathbf{Q}T^{-1}$ we can compute the induced distribution $P = QT^{-1}$ as

$$p(z,v,\Delta) = \{f(v-\theta z)\overline{H}(v\,|\,z)\}^{\Delta}\{h(v\,|\,z)\overline{F}(v-\theta z)\}^{1-\Delta}h(z)$$

(where $h(c|z)$ and $H(c|z)$ are the conditional density and distribution function of C conditional on $Z = z$, and $h(z)$ is the marginal density of Z), and differentiate as above. Or, alternatively, we can project the scores $\dot{\mathrm{l}}_i(\cdot,\mathrm{Q})$ in the model $\mathbf{Q}$ into scores in the model $\mathbf{P}$ by simply calculating their conditional expectation given the observation $T = T(X) = (Z, Y\wedge C, 1_{[Y\le C]}) = (Z, V, \Delta)$:

(5.6) $$\dot{\mathrm{l}}_i(T;\mathrm{P}) = E[\dot{\mathrm{l}}_i(X;\mathrm{Q})\,|\,T]\,,\qquad i = 1,2\,.$$

(For justifications of this conditional expectation formula, and for preservation of regular parametric models under measurable transformations more generally, see: Ibragimov and Has'minskii (1981), theorem 7.2, page 70; van der Vaart (1988), appendix A.3; and Le Cam and Yang (1988), section 7.) To calculate the conditional expectations in (5.6), first note that with

(5.7) $$\varepsilon \equiv Y - \theta Z \qquad \text{and} \qquad \delta \equiv C - \theta Z\,,$$

it follows that

(5.8) $$Y\wedge C - \theta Z = \varepsilon\wedge\delta \qquad \text{and} \qquad 1_{[Y\le C]} = 1_{[\varepsilon\le\delta]}\,,$$

so conditioning on T is equivalent to conditioning on $\tilde{T} \equiv (Z, \varepsilon\wedge\delta, 1_{[\varepsilon\le\delta]})$. Furthermore, note that ε and δ are independent. Thus the conditional expectation in (5.6) can be calculated as

$$E[\dot{\mathrm{l}}_1(X;\mathrm{Q})\,|\,Z, \mathrm{F}_t]$$

evaluated at $t = \delta$ where

$$\mathrm{F}_t \equiv \sigma\{\, 1_{[\varepsilon\le s]} : s \le t\,\} = \sigma\{\,\varepsilon\wedge t\,,\ 1_{[\varepsilon\le t]}\}\,.$$

But by proposition 3.1 it follows that

$$E[\dot{\mathbf{l}}_1(X;\mathbf{Q})|Z,\mathbf{F}_t] = E[Z\psi(\varepsilon)|Z,\mathbf{F}_t]$$

$$= Z\{1_{[\varepsilon \le t]}\psi(\varepsilon) + 1_{[\varepsilon > t]}\frac{E\,\psi(\varepsilon)1_{[\varepsilon > t]}}{1-F(t)}\} \tag{5.9}$$

$$= Z\{1_{[\varepsilon \le t]}\psi(\varepsilon) + 1_{[\varepsilon > t]}\frac{f}{1-F}(t)\} \tag{5.10}$$

$$= Z\int_{-\infty}^{t} R\psi\, d\mathbf{M} \tag{5.11}$$

where $M(t) \equiv 1_{[\varepsilon \le t]} - \int_{-\infty}^{t} 1_{[\varepsilon \ge s]}\, d\Lambda(s)$ and $\Lambda(t) = \int_{-\infty}^{t}(1-F)^{-1}\, dF$ is the cumulative hazard function corresponding to F. The third formula results from

$$E[\psi(\varepsilon)|\varepsilon > t] = \frac{f}{1-F}(t) = \lambda(t); \tag{5.12}$$

also note that

$$\begin{aligned} R\psi(t) &= \psi(t) - E[\psi(\varepsilon)|\varepsilon > t] \\ &= \psi(t) - \lambda(t) \\ &= -\frac{f'}{f}(t) - \frac{f}{1-F}(t)\,. \\ &= -\frac{d}{dt}\log\lambda(t) = -\frac{\lambda'(t)}{\lambda(t)}\,. \end{aligned} \tag{5.13}$$

Note that

$$\begin{aligned} J(t) &\equiv \int_{-\infty}^{t}(R\psi)^2\, dF \\ &= \int_{-\infty}^{t}(\frac{f'}{f})^2\, dF + \frac{f^2(t)}{1-F(t)} \\ &= \int_{-\infty}^{t}(\frac{\lambda'}{\lambda})^2\, dF \end{aligned}$$

from (3.8), (3.9), and (5.11); $J(t)$ is the information for location based on $\varepsilon \wedge t$.

Similarly, the score (operator) for f in the model $\mathbf{P}$ can be obtained by computing

$$E[\dot{l}_2 a(X;Q)\,|\,Z, \mathbf{F}_t] = E[a(\varepsilon)\,|\,Z, \mathbf{F}_t] = \int_{-\infty}^{t} Ra\, d\boldsymbol{M}\,. \tag{5.14}$$

Evaluating (5.11) and (5.14) at $t = \delta$ yields

$$\begin{aligned} \dot{l}_1(T;\mathbf{P}) &= Z\int_{-\infty}^{\delta} R\psi\, d\boldsymbol{M} \\ &= Z\int_{-\infty}^{\infty} 1_{[\delta \ge s]} R\psi(s)\, d\boldsymbol{M}(s) \\ &= Z\int_{-\infty}^{\infty} R\psi(s)\, d\boldsymbol{M}_{uc} \end{aligned} \tag{5.15}$$

and

$$\begin{aligned} \dot{l}_2 a(T;\mathbf{P}) &= \int_{-\infty}^{\delta} Ra\, d\boldsymbol{M} \\ &= \int_{-\infty}^{\infty} 1_{[\delta \ge s]} Ra(s)\, d\boldsymbol{M}(s) \\ &= \int_{-\infty}^{\infty} Ra\, d\boldsymbol{M}_{uc} \end{aligned} \tag{5.16}$$

where

$$\begin{aligned} \boldsymbol{M}_{uc}(t) &\equiv 1_{[\varepsilon\wedge\delta \le t,\, \Delta = 1]} - \int_{-\infty}^{t} 1_{[\varepsilon\wedge\delta \ge s]}\, d\Lambda(s) \\ &= 1_{[V - \theta^T Z \le t,\, \Delta = 1]} - \int_{-\infty}^{t} 1_{[V - \theta^T Z \ge s]}\, d\Lambda(s)\,. \end{aligned}$$

Thus for $a^* \in L_2^0(F)$

$$\dot{l}_1(T;\mathbf{P}) - \dot{l}_2 a^*(T;\mathbf{P}) = \int_{-\infty}^{\delta} \{ZR\psi - Ra^*\}\, d\boldsymbol{M}\,. \tag{5.17}$$

To find the efficient score function l_1^* for θ in the model $\mathbf{P}$, we want to find a function a^* with $\int a^* dF = 0$ so that

$$l_1^* \equiv \dot{l}_1 - \dot{l}_2 a^* \perp \dot{l}_2 a \qquad \text{in} \qquad L_2(P)$$

for all functions $a \in L_2^0(F)$; i.e.

$$E\{[\dot{l}_1 - \dot{l}_2 a^*]\,\dot{l}_2 a\} = 0 \qquad \text{for all} \qquad a \in L_2^0(F)\,. \tag{5.18}$$

This is just as in Begun, Hall, Huang, and Wellner (1983) except that here we are working in $L_2(P)$ rather than $L_2(\mu)$, and have replaced A by

$\dot{l}_2$, β^* by a^*, and β by a. But, by conditioning on Z and δ, the expectation in (5.18) is easily computed as the expectation of the predictable covariation process of the martingale transforms in (5.17) and (5.16): thus the left side of (5.18) equals

$$EE\{[\dot{l}_1 - \dot{l}_2 a^*]\dot{l}_2 a \mid Z, \delta\}$$

$$= EE\{\int_{-\infty}^{\infty} 1_{[\delta \geq s]}[ZR\psi(s) - Ra^*(s)]Ra(s)1_{[\varepsilon \geq s]}\, d\Lambda(s) \mid Z, \delta\}$$

$$= E\int_{-\infty}^{\infty} 1_{[\delta \geq s]}[ZR\psi(s) - Ra^*(s)]Ra(s)\, dF(s)$$

$$= \int_{-\infty}^{\infty} \{E[Z1_{[\delta \geq s]}]R\psi(s) - E[1_{[\delta \geq s]}]Ra^*\}Ra(s)\, dF(s)$$

$$= \int_{-\infty}^{\infty} L\{E[Z1_{[\delta \geq s]}]R\psi(s) - \overline{G}(s)Ra^*(s)\}a(s)\, dF(s), \tag{5.19}$$

by using $R^T = L$ from (2.10) where $\overline{G}(s) \equiv E1_{[\delta \geq s]} = P(\delta \geq s)$. From (5.19) it is easy to make the right choice of a^*: define

$$K(s) \equiv E[Z \mid \delta \geq s]$$

and set

$$a^* \equiv L(KR\psi) \equiv L(K(s)R\psi(s)).$$

Then, since $R \circ L$ = identity by proposition 2.1, it follows that

$$Ra^*(s) = K(s)R\psi(s) \tag{5.20}$$

and

$$\overline{G}(s)Ra^*(s) = E[Z1_{[\delta \geq s]}]R\psi(s), \tag{5.21}$$

so, with this choice of a^*, the right side of (5.19) is zero for all $a \in L_2^0(F)$. Thus, by substituting (5.20) in (5.17), (5.18) holds with the efficient score function

$$l_1^* \equiv \dot{l}_1 - \dot{l}_2 a^*$$

$$= \int_{-\infty}^{\infty} 1_{[\delta \geq s]}\{Z - E[Z \mid \delta \geq s]\}R\psi(s)\, d\mathbf{M}(s)$$

$$= \int_{-\infty}^{\infty} \{Z - E[Z \mid \delta \geq s]\}R\psi(s)\, d\mathbf{M}_{uc}(s)$$

$$= \int_{-\infty}^{\infty} \{Z - E[Z \mid V - \theta Z \geq s]\}R\psi(s)\, d\mathbf{M}_{uc}(s) \tag{5.22}$$

where the last line follows from independence of (Z, δ) and ε and (5.8). Thus the information for θ in the model $\mathbf{P}$ is, again by a martingale calculation,

$$
\begin{aligned}
I(\theta, \mathrm{P}) &= E\int_{-\infty}^{\infty} 1_{[\delta \geq s]}\{Z - E[Z \mid \delta \geq s]\}^2 (R\psi)^2(s)\, 1_{[\varepsilon \geq s]}\, d\Lambda(s) \\
&= E\int_{-\infty}^{\infty} 1_{[\delta \geq s]}\{Z - E[Z \mid \delta \geq s]\}^2 (R\psi)^2(s)\, dF(s) \\
&= \int_{-\infty}^{\infty} E\{ 1_{[\delta \geq s]}(Z - E(Z \mid \delta \geq s))^2 \} (R\psi)^2(s)\, dF(s) \qquad (5.23)\\
&= E[Z^2 J(C - \theta Z)] - \int K^2(s)\overline{G}(s)\, dJ(s).
\end{aligned}
$$

If Z and $\delta \equiv C - \theta Z$ are independent, then (5.23) reduces to

$$Var(Z)\int_{-\infty}^{\infty} P(\delta \geq s)(R\psi)^2(s)\, dF(s).$$

When there is no censoring, $P(\delta \geq s) = 1$ for all $s \geq 0$, and the information for θ reduces still further to the familiar expression

$$Var(Z) I_f,$$

where $I_f \equiv \int (f'/f)^2\, dF = \int (R\psi)^2\, dF$.

Taking $R\psi$ to be a given fixed function $R_0\psi_0$ ($R_0 \equiv R(F_0)$) in (5.22) yields the function

$$h(T) = \int_{-\infty}^{\infty} [Z - E(Z \mid \delta \geq s)]\, R\psi_0(s)\, d\mathbf{M}_{uc}(s),$$

which, by calculations exactly as those leading to (5.19), satisfies $h \perp \dot{\mathrm{l}}_2 a$ in $L_2(P)$ for all $a \in L_2^0(G)$. This suggests estimation of θ as a solution $\hat{\theta} = \hat{\theta}(\psi_0)$ of

$$
\begin{aligned}
0 &= \sum_{i=1}^{n} \int [Z_i - K(s; \hat{\theta})](R_0\psi_0)(s)\, dN_i(s) \\
&= \sum_{i=1}^{n} \Delta_i [Z_i - K(V_i - \hat{\theta}Z_i; \hat{\theta})] R_0\psi_0(V_i - \hat{\theta}Z_i)
\end{aligned}
$$

where

$$K(s\,;\theta) \equiv \frac{\sum_{i=1}^{n} Z_i \, 1_{[V_i - \theta Z_i \geq s]}}{\sum_{i=1}^{n} 1_{[V_i - \theta Z_i \geq s]}} .$$

This class of estimates, which are closely related to the rank estimators of Hodges and Lehmann (1963) for uncensored data, was suggested and studied by Tsiatis (1986) building on earlier work by Wei and Gail (1983) and Louis (1981). The particular choice $\psi_0(t) = e^t - 1$, corresponding to the extreme value distribution $1 - F_0(t) = \exp(-e^t)$ with hazard function $\lambda_0(t) = e^t$, has $R_0\psi_0(t) = -\lambda'_0(t)/\lambda_0(t) = -1$, and yields an estimator related to the log - rank test.

Acknowledgments. We owe thanks to Peter Bickel (for the proof of the boundedness of R), to Richard Gill (for the references to Chou and Meyer), and to Peter Sasieni (for pointing out the nice conditional expectation formula for the Cox model between (4.15) and (4.16)).

BIBLIOGRAPHY

1. Aalen, O. O., "Nonparametric inference for a family of counting processes," Ann. Statist., 6 (1978), 701 - 726.

2. Aalen, O. O., "A model for nonparametric regression analysis of counting processes," Lecture Notes in Statistics, 2, 1 - 25, Springer, New York, 1980.

3. Andersen, P. K. and Gill, R. D., "Cox's regression model for counting processes: a large sample study," Ann. Statist., 10 (1982), 1100-1120.

4. Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A., "Information and asymptotic efficiency in parametric and nonparametric models," Ann. Statist., 11 (1983), 432 - 452.

5. Beran, R., "Nonparametric regression with randomly censored survival data," Tech. Report., University of California, Berkeley, 1981.

6. Bickel, P. J., "Efficient testing in a class of transformation models," Papers on Semiparametric Models, MS-R8614 (1986), 63-81, Centrum voor Wiskunde en Informatica, Department of Mathematical Statistics, Amsterdam.

7. Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A., Efficient and Adaptive Inference in Semiparametric Models, Forthcoming Monograph, Johns Hopkins University Press, Baltimore, 1988.

8. Breslow, N. E. and Storer, B. E., "General relative risk functions for case - control studies," Am. J. Epidemiol., 122 (1985), 149 - 162.

9. Buckley, J. and James, I., "Linear regression with censored data," Biometrika, 66 (1979), 429-436.

10. Chou, C. S. and Meyer, P. A., "La representation des martingales relatives a un processus punctuel discret," C. R. Acad. Sci Paris, A 278 (1974), 1561 - 1563.

11. Chou, C. S. and Meyer, P. A. "Sur la representation des martingales comme integrales stochastiques dans les processus ponctuels," Seminaire de Probabilites IX, Lecture Notes in Mathematics 465 (1975), 226 - 236.

12. Clayton, D. and Cuzick, J., "The semiparametric Pareto model for regression analysis, of survival times," Papers on Semiparametric Models, MS-R8614 (1986), 19 - 31, Centrum voor Wiskunde en Informatica, Department of Mathematical Statistics, Amsterdam.

13. Cox, D. R., "Regression models and life tables (with discussion)," J. Roy. Statist. Soc. B 34 (1972), 187 - 220.

14. Cox, D. R. and Oakes, D., "Analysis of Survival Data," Chapman and Hall, London, 1984.

15. Csorgo, M., Csorgo, S., and Horvath, L., "An Asymptotic Theory for Empirical Reliability and Concentration Processes," Lecture Notes in Statistics, 33, Springer-Verlag, New York, 1986.

16. Dabrowska, D. M., "Nonparametric regression with censored survival time data," Scand. J. of Statistics, (1987), to appear.

17. Dabrowska, D. M., "Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates," Preprint, University of California, Berkeley, 1987.

18. Dunford, N. and Schwartz, J. T., Linear Operators, Part I, Interscience Publishers, J. Wiley and Sons, Inc., New York, 1958.

19. Efron, B., "The efficiency of Cox's likelihood function for censored data," J. Amer. Statist. Assoc., 72 (1977), 557 - 565.

20. Efron, B. and Johnstone, I., "Fisher's information in terms of the hazard rate," Preprint, Department of Statistics, Stanford, 1987.

21. Elliott, R. J., Stochastic Calculus and Applications, Springer-Verlag, New York, 1982.

22. Engle, R. F., Granger, C.W.J., Rice, J., and Weiss, A., "Semiparametric estimates of the relation between weather and electricity sales," J. Amer. Statist. Assoc., 81 (1986), 310 - 320.

23. Green, P., Jennison, C., and Seheult, A., "Analysis of field experiments by least squares smoothing," J. Roy. Statist. Soc., B 47 (1985), 299 - 314.

24. Green, P. and Yandell, B., "Semiparametric generalized linear models," Proceedings 2nd International GLIM Conference, Lecture Notes in Statistics, 32, Springer - Verlag, Berlin, 1985.

25. Green, P. J., "Penalized likelihood for general semiparametric regression models," Technical Report #2819, Department of Statistics, University of Wisconsin, Madison, 1985.

26. Hall, W. J. and Wellner, J. A., "Mean residual life," Statistics and Related Topics, 169 - 184, North Holland, Amsterdam, 1981, M. Csorgo, D. A. Dawson, J.N.K. Rao, and A.K.Md.E. Saleh, editors.

27. Hardy, G. H., Littlewood, J. E., and Polya, G., Inequalities, 2nd edition, Cambridge University Press, 1952.

28. Hastie, T. and Tibshirani, R., "Generalized additive models: some applications," J. Amer. Statist. Assoc., 82 (1987), 371 - 386.

29. Heckman, N. E., Spline smoothing in a partly linear model, J. Roy. Statist. Soc., B

48 (1986), 244 - 248.

30. Hodges, J. L. and Lehmann, E. L., "Estimates of location based on rank tests," Ann. Math. Statist., 34 (1963), 598 - 611.

31. Ibragimov, I. A. and Has'minskii, R. Z. Statistical Estimation: Asymptotic Theory. Springer-Verlag, New York, 1981.

32. James, I. and Smith, P. J., "Consistency results for linear regression with censored data," Ann. Statist., 12 (1984), 590 - 600.

33. Kalbfleisch, J. D. and Prentice, R. L. The Statistical Analysis of Failure Time Data. Wiley, New York, 1980.

34. Karlin, S. and Taylor, H. M., A First Course in Stochastic Processes, Second Edition, Academic Press, New York, 1975.

35. Koul, H., Susarla, V. and Van Ryzin, J. "Regression analysis with randomly right censored data," Ann. Statist. 9 (1981), 1276 - 1288.

36. Lawless, J. F. Statistical Models and Methods for Lifetime Data. Wiley, New York, 1982.

37. Le Cam, L. and Yang, G. L. "On the preservation of local asymptotic normality under information loss," Ann. Statist. 16 (1988), to appear.

38. Lipster, R. S. and Shiryayev, A. N. Statistics of Random Processes I and II. Springer-Verlag, New York, 1978.

39. Louis, T. A., "Non-parametric analysis of an accelerated failure time model," Biometrika, 68 (1981), 381 - 390.

40. McKeague, I. W., "Estimation for a semimartingale regression model using the method of sieves," Ann. Statist., 14 (1986), 579 - 589.

41. Miller, R. G., "Least squares regression with censored data," Biometrika, 63 (1976), 449 - 464.

42. Moolgavkar, S. and Venzon, D., "Confidence regions in curved exponential families: application to matched case-control and survival studies with general relative risk function," Ann. Statist., 15 (1987), 346 - 359.

43. O'Sullivan, Finbarr, "Relative risk estimation," Preprint, Univ. of California, Berkeley, 1986.

44. O'Sullivan, Finbarr, "Nonparametric estimation in the Cox proportional hazards model," Preprint, Univ. of California, Berkeley, 1986.

45. Oakes, D., "Aspects of partial likelihood," I.S.I. Review, 49 (1981), 235-264.

46. Prentice, R. L., "Linear rank tests with right censored data," Biometrika, 65 (1978), 165 - 179.

47. Prentice, R. L. and Self, S. G., "Asymptotic distribution theory for Cox-type regression models with general relative risk form," Ann. Statist., 11 (1983), 804-813.

48. Pyke, R., Spacings, J. Roy. Statist. Soc., B 27 (1965), 395 - 449.

49. Ritov, Y., "Efficient and unbiased estimation in nonparametric regression with censored data," Preprint, University of California, Berkeley, 1984.

50. Rudin, W., Functional Analysis, McGraw-Hill, New York, 1973.

51. Sasieni, P., Efficient estimation in semiparametric hazard models, Forthcoming Ph.D. dissertation, University of Washington, 1988.

52. Shorack, G. R. and Wellner, J. A., Empirical Processes with Applications to Statistics, Wiley, New York, 1986.

53. Thomas, D. C., "General relative risk models for survival time and matched case-control analysis," Biometrics, 37 (1981), 673 - 686, 1981.

54. Tsiatis, A. A., "Estimating regression parameters using linear rank tests for censored data," Preprint, Dept. of Biostatistics, Dana-Farber Cancer Institute, 1986.

55. van der Vaart, A., Statistical Estimation in Large Parameter Spaces. CWI Tract 44 (1988), Centrum voor Wiskunde en Informatica, Amsterdam.

56. Wei, L. J. and Gail, M. H., "Nonparametric estimation for a scale-change with censored observations," J. Amer. Statist. Assoc., 78 (1983), 382 -388.

57. Wellner, J. A., "Asymptotic optimality of the product limit estimator," Ann. Statist., 10, (1982), 595 - 602.

58. Zucker, D. M., Survival data regression analysis with time - dependent covariate effects, Unpublished Ph.D. Dissertation, Johns Hopkins University, 1986.

59. Zucker, D. M. and Karr, A. F., "Survival data regression analysis with time - dependent covariate effects: a penalized partial likelihood approach," Submitted to Ann. Statist., 1987.

DEPARTMENT OF STATISTICS
THE HEBREW UNIVERSITY OF JERUSALEM
MT. SCOPUS, JERUSALEM 91905
ISRAEL

DEPARTMENT OF STATISTICS GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA