The Annals of Statistics
1991, Vol. 19, No. 3, 1651–1660

AN E-ANCILLARITY PROJECTION PROPERTY OF COX'S PARTIAL SCORE FUNCTION[1]

BY I.-SHOU CHANG AND CHAO A. HSIUNG

National Central University, and Academia Sinica and Columbia University

This paper shows that Cox's partial score function is the projection of the score function on the (locally) E-ancillary subspace for the nuisance parameter (Small and McLeish). This is done by adapting the concepts of (locally) E-ancillarity and (locally) E-sufficiency for inference functions (McLeish and Small) to an extended Cox's regression model, where the baseline function is allowed to be a predictable process.

1. Introduction. In Small and McLeish (1988a) and McLeish and Small (1988), the concepts of ancillarity, sufficiency and completeness of statistics were extended to cover estimating functions or inference functions, which are applicable in a wider context than the standard notions. These generalized concepts were then used in Small and McLeish (1988b) to study problems involving the inferential separation of the data into informative and noninformative components. In particular, they explored a projection method for eliminating the nuisance parameter, which reduces to conditioning on the complete sufficient statistic for this parameter if the complete sufficient statistic exists.

We recapitulate here some of their basic ideas. Consider a two parameter inference function $\phi(\theta, \lambda; x)$ which is unbiased in the sense that $E_{(\theta, \lambda)}\phi(\theta, \lambda; x) = 0$. Fix θ for the moment. In the resulting one-parameter model we can decompose the function ϕ into its sufficient and ancillary components with respect to the parameter λ: $\phi = \phi_s + \phi_a$. When λ is a nuisance parameter, the appropriate inference function for the problem should be insensitive to λ. Thus attention naturally turns toward ϕ_a. It is to be hoped that if ϕ is chosen so as to be sensitive to the parameter of interest θ, then the resulting ϕ_a will possess this property and in addition be insensitive to the nuisance parameter. Roughly speaking, the mapping from ϕ to ϕ_a is the projection under study.

Among many interesting ideas and examples, Small and McLeish (1988b) discussed Cox's proportional hazards model in order to explain this projection method as a technique to reduce sensitivity of an inference function with respect to nuisance parameters. More precisely, with some spade work, it is conjectured there that Cox's partial score function is the projection of the

Received July 1989; revised July 1990.
[1]Research partially supported by NSC of ROC.
AMS 1980 *subject classifications*. Primary 62A99, 62B99; secondary 62J99, 62M99.
Key words and phrases. Inference functions, score function, ancillarity, sufficiency, completeness, nuisance parameter, Cox's partial score function.

score function on the E-ancillary subspace for the nuisance parameter. The purpose of this paper is to establish the validity of this conjecture in modified forms, which provides a new motivation for the use of maximum partial likelihood estimation in Cox's regression model [cf. Chang and Hsiung (1990)].

The plan of this paper is as follows. Section 2 fixes the notation and the model assumptions. In fact, we consider a more general Cox's regression model for counting processes, where we allow the baseline function to be a predictable process, instead of only a deterministic function. Section 3 adapts the extended concepts of ancillarity, sufficiency and completeness to this model. Section 4 gives the main theorems and their proofs, which contains two subsections dealing with a local version and a global version of this projection property, respectively. Although this paper is self-contained, we refer the readers to Small and McLeish (1988b, 1989) for more background and related concepts.

Finally, we would like to remark that Small and McLeish (1988b) found that the locally E-ancillary projection of the score function for Cox's model with deterministic baseline function is similar to but not equal to the partial score function. In some sense, the success of our work hinges on the enlargement of the nuisance parameter space, which is crucial in obtaining (4.4) in the local case and in obtaining Lemma 4.1 in the global case. This enlargement of the nuisance parameter space to include predictable processes is also desirable from practical considerations. More detailed discussions with an example in industrial context is given in Section 2.

2. Notation and the model. This section fixes the notation and the model for the discussion in this paper. It also contains a likelihood ratio formula to be used later.

Let $N(t) = (N_1(t), \dots, N_K(t))$, $t \geq 0$, be a K-variate counting process. Assume that, relative to a filtration F_t, $N(t)$ has intensity $\tilde{\lambda}(t) = (\lambda_1(t), \dots, \lambda_K(t))$ of the form

$$\lambda_k(t) = \lambda(t) Y_k(t) r(\theta' Z_k(t)), \tag{2.1}$$

where $r(\cdot) \geq 0$ is a known twice differentiable function, $\theta \in \Theta \subset R^d$, $Y_k(\cdot) \geq 0$ is a bounded predictable process and $Z_k(\cdot)$ is a R^d-valued bounded predictable process. When $\lambda(\cdot) \geq 0$ is an unknown deterministic function, (2.1) describes the Cox regression model for counting processes, which were studied by Andersen and Gill (1982), Prentice and Self (1983) and many others.

In this paper, we shall consider (2.1) with relaxed condition on the baseline function λ. For a nonnegative process λ, we call $\{t \geq 0 | \lambda(t) > 0 \text{ a.e.}\}$ the proper support of λ. Let Λ be the set of all nonnegative predictable processes bounded on every compact subset of $[0, \infty)$ with proper support a given set C. The model (2.1) to be considered in this paper assumes that $\lambda \in \Lambda$.

The statistical problem we have is to estimate θ based on the data

$$\{N_k(t), Y_k(t), Z_k(t) | k = 1, \dots, K, 0 \leq t \leq T\}$$

at some stopping time T, treating $\lambda \in \Lambda$ as a nuisance parameter. In the model for survival data, $N_k(t)$ and $Y_k(t)$ together indicate the status of the kth subject at time t and $Z_k(t)$ denotes the covariate of the kth subject at time t.

It may seem a little unorthodox to use a predictable process as a parameter in a statistical problem. However, we would like to point out that multivariate failure time data with nondeterministic baseline function $\lambda(\cdot)$ did appear from practical considerations [see, e.g., pages 373–375 of Prentice, Williams and Petersen (1981)]. A simple example of a (nondeterministic) predictable baseline process is

$$\lambda(t) = \sum_{i=0}^{\infty} h_i(t - \tau_i) 1_{(\tau_i, \tau_{i+1}]}(t),$$

where $\tau_i = \inf\{t > 0 | \sum_{k=1}^{K} N_k(t) = i\}$ and h_i is a deterministic function. In industrial context, this is the situation that components of a machine share a common hazard rate function h_i, which depends on the total number of events experienced by the machine.

On the theoretical side, we know from the direct and converse Radon–Nikodym derivative theorems for multivariate point processes [see, e.g., Brémaud (1981), pages 166, 168, 187, 242; Gill (1980), page 14] that nonnegative predictable processes provide a very effective way to parametrize the probability measures on the sample path space. It is based on these theorems that we are able to give some properties of the likelihood function on model (2.1).

It follows from the direct Radon–Nikodym derivative theorem that every parameter value $(\theta, \lambda) \in \Theta \times \Lambda$ specifies a probability measure $P^{(\theta, \lambda)}$ on the sample path space. Denote by $P_T^{(\theta, \lambda)}$ the restriction of $P^{(\theta, \lambda)}$ to F_T, where F_T is the σ-field of events up to a stopping time T. We will use $E_{(\theta, \lambda)}$ to denote the expectation corresponding to the probability measure $P^{(\theta, \lambda)}$.

Suppose $0 \in \Theta$, $r(0) = 1$. Let $\lambda_0(t) = I_C(t)$. Let $T_k(n) = \inf\{t \geq 0 | N_k(t) \geq n\}$. Then, the Radon–Nikodym derivative with respect to the dominating measure specified by $(0, \lambda_0)$ is

$$\frac{dP_T^{(\theta, \lambda)}}{dP_T^{(0, \lambda_0)}} = L(T, \theta, \lambda) = \prod_{k=1}^{K} L_k(T, \theta, \lambda), \tag{2.2}$$

where

$$L_k(T, \theta, \lambda) = \left(\prod_{n=1}^{\infty} \lambda(T_k(n)) r(\theta' Z_k(T_k(n))) I_{[T_k(n) \leq T]} \right) \times \exp \int_0^T (1 - \lambda(s) r(\theta' Z_k(s))) Y_k(s) \lambda_0(s)\, ds.$$

It is clear that

$$\begin{aligned} \log L_k(T, \theta, \lambda) &= \int_0^T \log(\lambda(s) r(\theta' Z_k(s)))\, dN_k(s) \\ &\quad + \int_0^T (1 - \lambda(s) r(\theta' Z_k(s))) Y_k(s) \lambda_0(s)\, ds. \end{aligned} \tag{2.3}$$

Let λ and α be two predictable processes. Assume that both λ and $\lambda + \varepsilon\alpha$ belong to Λ for all small enough positive ε. Let

$$M_k(t) = N_k(t) - \int_0^t \lambda_k(s)\, ds$$

be the basic martingale. Let $M(t) = \sum_{k=1}^K M_k(t)$. Then, using (2.2) and (2.3), we get by straightforward calculations that

$$\frac{L(T,\theta,\lambda+\varepsilon\alpha)}{L(T,\theta,\lambda)} = 1 + \varepsilon \int_0^T \frac{\alpha(s)}{\lambda(s)}\, dM(s) + O(\varepsilon^2). \tag{2.4}$$

We note that $\alpha(s)/\lambda(s)$ is interpretated as 0 whenever $\lambda(s)$ is 0.

We obtain also from (2.2) and (2.3) that the score function in θ with fixed λ is

$$\begin{aligned} \frac{\partial}{\partial\theta_l} \log L(T,\theta,\lambda) = {} & \sum_{k=1}^K \int_0^T \frac{r^{(1)}(\theta' Z_k(s))}{r(\theta' Z_k(s))} Z_{k,l}(s)\, dN_k(s) \\ & - \sum_{k=1}^K \int_0^T r^{(1)}(\theta' Z_k(s)) Z_{k,l}(s) Y_k(s) \lambda(s)\, ds, \end{aligned} \tag{2.5}$$

where $r^{(1)}(\cdot)$ is the derivative of r, $Z_{k,l}$ is the lth component of Z_k.

3. *E*-ancillarity and *E*-sufficiency. Motivated by the classical Basu's theorem on ancillarity and sufficiency, Small and McLeish (1988a, b) define the concepts of E-ancillarity and E-sufficiency for inference functions and some local versions of them. In this section, we shall adapt these concepts to the Cox's regression model of Section 2, which enables us to describe the projection method that desensitizes the score function with respect to the nuisance parameter.

The class Ψ of inference functions we shall consider consists of functions of the form

$$\psi(T,\theta,\lambda) = \sum_{k=1}^K \int_0^T f_k(t,\theta,\lambda)\, dM_k(t), \tag{3.1}$$

where $f_k(t,\theta,\lambda)$ is a predictable process so that $\psi(t,\theta,\lambda)$ is a square-integrable martingale for every $(\theta,\lambda) \in \Theta \times \Lambda$. Intuitively speaking, we make the natural requirement that the inference function has mean zero for every stopping time. Chang and Hsiung (1990) contains some discussion in this regard. Small and McLeish (1988b) also considers inference functions of this form.

An element $\phi \in \Psi$ is called an E-ancillary function in λ if

$$E_{(\theta,\eta)}\phi(T,\theta,\lambda) = 0 \tag{3.2}$$

for every $\theta \in \Theta$, $\eta, \lambda \in \Lambda$. In fact, we will assume in this paper that

$$E_{(\theta,\lambda)}\left(\frac{L(T,\theta,\eta)}{L(T,\theta,\lambda)}\right)^2 < \infty \tag{3.3}$$

for every $\theta \in \Theta$, $\eta, \lambda \in \Lambda$ when (global) E-ancillarity and related concepts are in consideration.

An element $\phi \in \Psi$ is called a locally E-ancillary function in λ if

$$E_{(\theta, \lambda + \varepsilon\alpha)}\phi(T, \theta, \lambda) = o(\varepsilon) \tag{3.4}$$

as $\varepsilon \downarrow 0$, whenever $\theta \in \Theta$, $\lambda, \lambda + \varepsilon\alpha \in \Lambda$ and

$$\begin{aligned} &E_{(\theta, \lambda)} \exp \varepsilon \int_0^T \frac{\alpha(s)}{\lambda(s)} \, dM_k(s) < \infty, \\ &E_{(\theta, \lambda)} \exp \varepsilon \int_0^T \gamma\left(\frac{\alpha(s)}{\lambda(s)}\right) dN_k(s) < \infty \end{aligned} \tag{3.5}$$

for every small enough positive ε, every $(\theta, \lambda) \in \Theta \times \Lambda$, every $k = 1, \dots, K$. Here γ is the function satisfying $\log(1 + ax) - ax = a^2\gamma(x)$. The moment condition (3.5) will be used in the derivation of (3.9).

Let $A(Al)$ denote the set of all (locally) E-ancillary functions in λ. Let

$$\begin{aligned} A'(Al') = \{\phi \in \Psi \,|\, &E_{(\theta, \lambda)}(\phi(T, \theta, \lambda) - \phi_n(T, \theta, \lambda))^2 \text{ goes to } 0 \\ &\text{for every } (\theta, \lambda) \in \Theta \times \Lambda \text{ for some sequence } \phi_n \in A(Al)\}. \end{aligned} \tag{3.6}$$

We will call $A'(Al')$ the space of (locally) E-ancillary functions in λ.

With $A'(Al')$, we are able to define the corresponding concepts of (locally) E-sufficiency and completeness as follows.

A subset $S(Sl)$ of Ψ is called (locally) E-sufficient in λ if $\phi \in \Psi$ together with

$$E_{(\theta, \lambda)}\phi(T, \theta, \lambda)\psi(T, \theta, \lambda) = 0$$

for every $\psi \in S(Sl)$, every $(\theta, \lambda) \in \Theta \times \Lambda$ implies $\phi \in A'(Al')$. It is called complete (locally) E-sufficient in λ if the condition $\phi \in A'(Al')$ is also sufficient.

Let

$$\begin{aligned} Sg = \Big\{\psi \,|\, &\psi(T, \theta, \lambda) \text{ is a finite linear combination of elements} \\ &\text{in } \Psi \text{ of the form } \frac{L(T, \theta, \eta)}{L(T, \theta, \lambda)} - 1, \eta \in \Lambda\Big\}. \end{aligned} \tag{3.7}$$

Let Sc be the closure of Sg in the sense that A' is the closure of A.

It follows from the identity

$$E_{(\theta, \eta)}\phi(T, \theta, \lambda) = E_{(\theta, \lambda)}\phi(T, \theta, \lambda)\left(\frac{L(T, \theta, \eta)}{L(T, \theta, \lambda)} - 1\right)$$

for every $\phi \in \Psi$, that Sg is E-sufficient in λ. A little calculation using the definition of Sc, Schwarz inequality and (3.3) gives the following proposition.

PROPOSITION 3.1. *Sc is the complete E-sufficient space in* λ.

The rest of this section attempts to characterize the elements of the complete locally E-sufficient space in λ.

Using (2.4), we get

$$\begin{aligned} &E_{(\theta,\lambda+\varepsilon\alpha)}\phi(T,\theta,\lambda) \\ &\quad = E_{(\theta,\lambda)}\phi(T,\theta,\lambda)\frac{L(T,\theta,\lambda+\varepsilon\alpha)}{L(T,\theta,\lambda)} \\ &\quad = E_{(\theta,\lambda)}\phi(T,\theta,\lambda)\left(1+\varepsilon\int_0^T\frac{\alpha(s)}{\lambda(s)}\,dM(s)+\varepsilon^2R(T,\varepsilon)\right), \end{aligned} \tag{3.8}$$

where

$$R(T,\varepsilon)=\varepsilon^{-2}\left(\frac{L(T,\theta,\lambda+\varepsilon\alpha)}{L(T,\theta,\lambda)}-1-\varepsilon\int_0^T\frac{\alpha(s)}{\lambda(s)}\,dM(s)\right).$$

It follows from (3.5) that every term in (3.8) has second moment. Using (3.5), we can also get the uniform integrability of $\phi(T,\theta,\lambda)R(T,\varepsilon)$, which implies

$$E_{(\theta,\lambda+\varepsilon\alpha)}\phi(T,\theta,\lambda)=\varepsilon E_{(\theta,\lambda)}\phi(T,\theta,\lambda)\int_0^T\frac{\alpha(s)}{\lambda(s)}\,dM(s)+O(\varepsilon^2), \tag{3.9}$$

if $\phi\in\Psi$. Hence (3.4) holds if and only if

$$E_{(\theta,\lambda)}\phi(T,\theta,\lambda)\int_0^T(\alpha(s)/\lambda(s))\,dM(s)=0.$$

Therefore, we have proved the following proposition.

PROPOSITION 3.2. *The complete locally E-sufficient set Slc in* λ *is the space of integrals in* Ψ *that take the form* $\int_0^T g(s)\,dM(s)$, *with* $\lambda, \alpha = g\lambda$ *satisfying* (3.5).

4. The projection property of Cox's partial score function. We are ready to show that Cox's partial score function is the projection of the score function (2.5) into the space $A'(Al')$ of (locally) E-ancillary functions in λ, which desensitizes the score function with respect to the nuisance parameter λ as described in Small and McLeish (1988b).

After a few preliminaries, we will present the local version and the global version of this projection property in Section 4.1 and 4.2, respectively.

Let

$$J_l(s)=\frac{\sum_{k=1}^K r^{(1)}(\theta' Z_k(s))Y_k(s)Z_{k,l}(s)}{\sum_{k=1}^K r(\theta' Z_k(s))Y_k(s)} \tag{4.1}$$

$$G_l(T,\theta,\lambda)=\sum_{k=1}^K\int_0^T\left(\frac{r^{(1)}(\theta' Z_k(s))}{r(\theta' Z_k(s))}Z_{k,l}(s)-J_l(s)\right)dM_k(s), \tag{4.2}$$

$$U_l(T,\theta,\lambda)=\sum_{k=1}^K\int_0^T J_l(s)\,dM_k(s). \tag{4.3}$$

We note that $G_l(T,\theta,\lambda) = G_l(T,\theta)$ is, in fact, independent of λ and is the lth component of Cox's partial score function [cf. Andersen and Gill (1982), Prentice and Self (1983)].

Assume that both $G_l(\cdot,\theta,\lambda)$ and $U_l(\cdot,\theta,\lambda)$ are square-integrable martingales. These are mainly moment conditions, since both of them are stochastic integrals with martingale integrators and predictable integrands [cf. Chang and Hsiung (1990), (4.1) and (4.2)]. Thus both $G_l(T,\theta)$ and $U_l(T,\theta,\lambda)$ are in Ψ.

4.1. *Local version.* With a little condition on r, Y_k, Z_k, it follows immediately from Proposition 3.2 that

$$U_l(T,\theta,\lambda) \in Slc. \tag{4.4}$$

THEOREM 4.1.

$$\frac{\partial}{\partial\theta_l}\log L(T,\theta,\lambda) = G_l(T,\theta) + U_l(T,\theta,\lambda) \tag{4.5}$$

with $G_l \in Al'$, $U_l \in Slc$.

PROOF. It is obvious from (2.5) that (4.5) holds. By (4.4), it remains to show $G_l \in Al$ or $E_{(\theta,\lambda)}G_l(T,\theta)\phi(T,\theta,\lambda) = 0$ for every $\phi \in Slc$, $(\theta,\lambda) \in \Theta \times \Lambda$. This is obtained by calculating the predictable mutual variation of the square-integrable martingales $G_l(t)$ and $\phi(t,\theta,\lambda) = \sum_{k=1}^K \int_0^t g(s)\,dM_k(s)$.

Observe

$$\begin{aligned}
&\langle G_l(\cdot), \phi(\cdot,\theta,\lambda)\rangle_t \\
&\quad = \sum_{k=1}^K \left\langle \int_0^{\cdot} \left(\frac{r^{(1)}(\theta' Z_k(s))}{r(\theta' Z_k(s))} Z_{k,l}(s) - J_l(s) \right) dM_k(s), \int_0^{\cdot} g(s)\,dM_k(s) \right\rangle_t \\
&\quad = \sum_{k=1}^K \int_0^t g(s) \left\{ \frac{r^{(1)}(\theta' Z_k(s))}{r(\theta' Z_k(s))} Z_{k,l}(s) - J_l(s) \right\} r(\theta' Z_k(s)) Y_k(s) \lambda(s)\,ds \\
&\quad = 0.
\end{aligned}$$

This completes the proof. In fact, similar arguments imply

$$E_{(\theta,\eta)}G_l(T,\theta,\lambda) = 0,$$

which gives an alternative proof. □

4.2. *Global version.* Although the global version takes the same form as its local counterpart, its proof is more involved. We shall assume some conditions in order to ease the presentation. These conditions are by no means necessary. There are other sets of sufficient conditions, although we treat here one of the most important cases.

Assume the covariates Z_k's have nonnegative components. Assume $J_l \geq 0$ is bounded. Assume

$$\int_0^\infty \lambda(s)\,ds < \tilde{C}_0 \tag{4.6}$$

for some constant $\tilde{C}_0$. Then we have the following theorem.

THEOREM 4.2.

$$\frac{\partial}{\partial \theta_l} \log L(T,\theta,\lambda) = G_l(T,\theta) + U_l(T,\theta,\lambda), \tag{4.7}$$

with $G_l \in A$, $U_l \in Sc$.

Since (4.7) is obvious from (2.5) and $G_l \in A$ can be shown by straightforward calculation, we need only to show $U_l \in Sc$, which follows from Lemma 4.1 and Proposition 3.1.

LEMMA 4.1. $U_l(T,\theta,\lambda) \in Sg$, *defined in* (3.7).

PROOF. It suffices to exhibit an element $\eta \in \Lambda$ so that

$$\frac{L(T,\theta,\eta)}{L(T,\theta,\lambda)} - 1 = U_l(T,\theta,\lambda). \tag{4.8}$$

According to (2.2), (2.3) and the exponential formula for martingales [see, e.g., Brémaud (1981), page 166, (2.4)],

$$\frac{L(t,\theta,\eta)}{L(t,\theta,\lambda)} - 1 = \sum_{k=1}^{K} \int_0^t \frac{L(s-,\theta,\eta)}{L(s-,\theta,\lambda)}(\mu(s)-1)\,dM_k(s), \tag{4.9}$$

where $\mu = \eta/\lambda$. Therefore, it suffices to find a nonnegative predictable process μ satisfying

$$\frac{L(t-,\theta,\eta)}{L(t-,\theta,\lambda)}(\mu(t)-1) = J_l(t),$$

where J_l is defined in (4.1); or equivalently, by the likelihood ratio formula (2.2) and (2.3),

$$\left\{\exp\left\{\sum_{k=1}^{K}\int_0^{t-}\log\mu(s)\,dN_k(s) + \int_0^t (1-\mu(s))\left(\sum_{k=1}^{K}\lambda_k(s)\right)ds\right\}\right\}(\mu(t)-1) = J_l(t). \tag{4.10}$$

Since (4.10) is a first order ordinary differential equation on each stochastic interval $(\tau_j, \tau_{j+1}]$, where $\tau_j = \inf\{t > 0 | \sum_{k=1}^{K} N_k(t) = j\}$, we will only indicate its solution and omit details.

(i) On $[0, \tau_1)$, set

$$\mu(t) = 1 + \frac{J_l(t)}{C_0 - \int_0^t J_l(s)\left(\sum_{k=1}^{K}\lambda_k(s)\right) ds}, \tag{4.11}$$

where C_0 is a large constant such that the denominator in (4.11) is positive.

(ii) At τ_1, set

$$\mu(\tau_1) = 1 + \left\{\exp\int_0^{\tau_1}(\mu(s) - 1)\left(\sum_{k=1}^{K}\lambda_k(s)\right) ds\right\} J_l(\tau_1). \tag{4.12}$$

(iii) Suppose we have defined bounded predictable process $\mu \geq 1$ on $[0, \tau_j]$ satisfying (4.10). On (τ_j, τ_{j+1}), set

$$\mu(t) = 1 + A_j^{-1}\frac{J_l(t)}{C_j - A_j^{-1}\int_{\tau_j}^t J_l(s)\left(\sum_{k=1}^{K}\lambda_k(s)\right) ds} \tag{4.13}$$

for some suitably large constant C_j, where

$$A_j = \prod_{h=1}^{j}\mu(\tau_h)\exp\int_0^{\tau_j}(1 - \mu(s))\left(\sum_{k=1}^{K}\lambda_k(s)\right) ds.$$

The definition of μ at τ_{j+1} can be done similarly as that at τ_1 in (ii).

Thus, with the predictable process μ satisfying (4.10), the proof of this lemma is complete. □

Acknowledgments. Thanks are due to Professors D. Siegmund and K. C. Li for helpful conversations during the preparation of this paper. The second author would also like to thank Professor McLeish for background information when preparing this revision. We are also grateful for the suggestions and comments made by the referees.

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

BRÉMAUD, P. (1981). *Point Processes and Queues, Martingale Dynamics*. Springer, New York.

CHANG, I.-S. and HSIUNG, C. A. (1990). Finite sample optimality of maximum partial likelihood estimation in Cox's model for counting processes. *J. Statist. Plann. Inference* **25** 35–42.

GILL, R. D. (1980). *Censoring and Stochastic Integrals*. Math. Centre Tract **124**. Mathematisch Centrum, Amsterdam.

MCLEISH, D. L. and SMALL, C. G. (1988). *The Theory and Applications of Statistical Inference Functions. Lecture Notes in Statist.* **44**. Springer, New York.

PRENTICE, R. L. and SELF, S. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

PRENTICE, R. L., WILLIAMS, B. J. and PETERSEN, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68** 373–379.

SMALL, C. G. and MCLEISH, D. L. (1988a). Generalizations of ancillarity, completeness and sufficiency in an inference function space. *Ann. Statist.* **16** 534–551.

SMALL, C. G. and MCLEISH, D. L. (1988b). Projection as a method for increasing sensitivity and eliminating nuisance parameters. Statistics Technical Report Series, STAT-88-11, Dept. Statistics and Actuarial Science, Univ. Waterloo.

SMALL, C. G. and MCLEISH D. L. (1989). Projection as a method for increasing sensitivity and eliminating nuisance parameters. *Biometrika* **76** 693–703.

DEPARTMENT OF MATHEMATICS
NATIONAL CENTRAL UNIVERSITY
CHUNGLI, TAIWAN
REPUBLIC OF CHINA

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA