The Annals of Statistics
1994, Vol. 22, No. 3, 1275–1298

INFORMATION AND ASYMPTOTIC EFFICIENCY IN SOME GENERALIZED PROPORTIONAL HAZARDS MODELS FOR COUNTING PROCESSES[1]

By I-Shou Chang and Chao A. Hsiung

National Central University and Academia Sinica

Proportional hazards models with stochastic baseline hazards and estimators of the relative risk coefficient in these models were proposed by Prentice, Williams and Peterson and by Chang and Hsiung in medical and industrial contexts. The form of the estimating functions recommended varies according to the form of the unknown stochastic baseline hazards. This paper examines the same estimation problem in the context of large-sample theory. It is shown that the proposed estimators are regular, asymptotically normal and asmptotically efficient. Asymptotic information and representation theorems in the sense of Begun, Hall, Huang and Wellner are also provided for these models.

1. Introduction and motivations.

1.1. *Generalized proportional hazards models.* The proportional hazards model of survival analysis and its analysis by the method of partial likelihood originate in the work of Cox (1972, 1975). Since its introduction, it has been at the center of many important statistical developments. In particular, Andersen and Gill (1982) formulated Cox's regression model for counting processes and studied multivariate failure time data using martingale methods.

To be precise, Andersen and Gill (1982) considered K-variate counting process $N_1(t) = (N_{11}(t), \dots, N_{1K}(t))$ for which the intensity of $N_{1k}(t)$, relative to a filtration of $N_1(t)$, has the form

$$\lambda_{1k}(t) = \lambda_{10}(t) Y_{1k}(t) \exp\left[\theta Z_{1k}(t)\right], \tag{1.1}$$

where $\lambda_{10}(\cdot) \geq 0$ is a deterministic baseline hazard rate function, $Y_{1k}(\cdot) \geq 0$ and $Z_{1k}(\cdot)$ are bounded predictable processes and $\theta \in \Theta \subset R$ is the relative risk regression coefficient.

While the counting process formulation (1.1) of Andersen and Gill (1982) significantly enlarged the domain of application of proportional hazards models and the related concepts of Cox (1972, 1975), Prentice, Williams and Peterson (1981) indicated that, in medical research and in industry, there are proportional hazards models for multivariate failure time data which do not fall in the realm of (1.1). The situation is that the baseline hazard rate λ_{10} may be a

Received October 1990; revised September 1993.
[1]Research partially supported by National Science Council of the Republic of China.
AMS 1991 *subject classifications.* Primary 62M99; secondary 62J99, 62E99.
Key words and phrases. Asymptotic information, effective score, martingale, semiparametric model.

random process depending on the past history of the counting process, and the hazard rates for different values of θ are still proportional.

A simple example with random baseline hazard rate in the notation of (1.1) is described by setting

$$\lambda_{10}(t) = \sum_{i=0}^{\infty} h_i(t - T_{1i}) 1_{(T_{1i}, T_{1i+1}]}(t), \tag{1.2}$$

where $T_{1i} = \inf\{t > 0 \mid \sum_{k=1}^{K} N_{1k}(t) = i\}$, and $h_i(\cdot) \geq 0$ is a deterministic function.

An important special case of (1.2) is

$$\lambda_{10}(t) = \sum_{i=0}^{\infty} h_i 1_{(T_{1i}, T_{1i+1}]}(t), \tag{1.3}$$

with h_i being constants.

Just as in Andersen and Gill (1982), the statistical problem we are interested in for these and some other generalized proportional hazards models is to estimate the relative risk regression coefficient θ based on $N_1, \dots, N_J$ and related observables, treating h_i as nuisance prameters. Here $N_1, \dots, N_J$ are independent and indentically distributed K-variate counting processes with J being the number of study subjects and K the number of different failure-type classes for each subject, which implies that they have common parameters θ and h_i.

Chang and Hsiung (1991a) and Prentice, Williams and Peterson (1981) proposed estimators for this problem. The form of the estimating functions they proposed varies according to the size of the infinite-dimensional nuisance parameter space, the number of replicates observed and the univariateness of the counting processes. In this paper, we will focus on some large-sample properties of these estimators and establish their asymptotic efficiency.

At this point, we would like to remark that models (1.2) and (1.3) represent an enlargement of the nuisance parameter space in (1.1). One desirability of this kind of enlargement from the theoretical perspective was explained in Chang and Hsiung (1991b), where an E-ancillarity projection property of Cox's partial score function was established. We note that an analogous enlargement of parameter space in the multiplicative intensity model for counting processes was discussed in Millar (1989), which leads to the development of a general local asymptotic minimax theory.

In the rest of this subsection, we will indicate the relevance of (1.2) in applications by two examples. In an industrial context, we consider a machine with K components. Let $N_{1k}(t)$ denote the number of breakdowns of the kth component of the machine up to time t. Then (1.2) describes the situation that components of the machine share a common baseline hazard rate which depends only on the total number of breakdowns of the machine and the time from last breakdown.

In a medical context, a two-sample special case of (1.2) studied by Gail, Santner and Brown (1980) is described as follows. Assume $K = 1$ and $Z_{11}(t) = 0$ or 1 for every $t \geq 0$. Let $N_{j1}(t)$ denote the number of certain tumor occurrences of the jth patient up to time t in an experiment with J patients in total. Then (1.2) means that patients having had the same number of tumor occurrences are in a stratum assuming a common baseline hazard rate function for next occurrence. As usual, the true value of θ is a description of the treatment effect.

For more examples, we refer the readers to Prentice, Williams and Peterson [(1981), Section 2] and the references therein.

1.2. *Main results.* Our purpose in this paper is to examine these generalized proportional hazards models from the perspective of large-sample theory. Our study of large-sample theory starts with the information calculations, based on which the effective score functions can be obtained. In fact, the effective score functions are the limit of the estimating functions proposed by Chang and Hsiung (1991a) and Prentice, Williams and Peterson (1981). We then establish asymptotic normality and regularity of the resulting estimators. This together with information calculations and convolution theorems establishes the asymptotic efficiency of these estimators in their appropriate models.

In this paper, we will consider four types of generalized proportional hazards models. Model 1 is the well-known Cox model. Model 2 relates to (1.3), and Model 3 is (1.2). Model 4 includes both Models 2 and 3 as special cases.

We will see that the estimating functions proposed for these four models are all different. In particular, the proposed estimating function in Model 4 is identical with its "effective score function," and this is not the case in any other model.

It seems that the underlying idea in our approach, including the derivation of (optimal) estimating function, information calculations and the proof that the resulting estimators are regular, is to exhibit certain structures of orthogonal martingales, namely, martingales independent of the nuisance parameter and its orthogonal complements. The idea is carried out by considering martingales relative to the natural calendar time filtration in Models 1, 2 and 4. However, we need to consider martingales relative to different filtrations to make this idea rigorous in Model 3.

This paper is organized as follows. Section 2 fixes the notation and describes generalized proportional hazards models. Section 3 studies the information calculations and explains how to obtain the effective score functions or maximum partial likelihood (MPL) equations. Section 4 presents the asymptotic normality of these resulting estimators. Section 5 gives local asymptotic normality and convolution theorems and proves that MPL equations suggested in Section 3 lead to regular and, hence, asymptotically efficient estimators. Martingales, stochastic integrals and related concepts used freely in this paper can be found, for example, in Brémaud (1981), Gill (1980) and Elliott (1982).

2. Notation and the models. This section gives the notation and model assumptions which are used throughout the paper.

Let $N_1, \dots, N_J, \dots$ be a sequence of independent and identically distributed K-variate counting processes. Let $\mathcal{F}_{j,t}$ be the self-excited filtration of N_j, namely, $\mathcal{F}_{j,t}$ is the σ-field generated by $\{N_{jk}(s) \mid 0 \le s \le t, k = 1, \dots, K\}$. Assume that, relative to $\mathcal{F}_{j,t}$, the counting process $N_j(t) = (N_{j1}(t), \dots, N_{jK}(t))$ has intensity $\lambda_j(t) = (\lambda_{j1}(t), \dots, \lambda_{jK}(t))$ of the form

$$\lambda_{jk}(t) = \lambda_{j0}(t) Y_{jk}(t) \exp\left[\theta Z_{jk}(t)\right], \tag{2.1}$$

where $\lambda_{j0}(\cdot) \geq 0, Y_{jk}(\cdot) \geq 0$ and $Z_{jk}(\cdot)$ are bounded predictable processes defined for $t \in [0, \infty)$, and $\theta \in \Theta \in \mathbb{R}$ is the relative risk regression coefficient. We assume that Θ is a bounded set. Let $M_{jk}(t) = N_{jk}(t) - \int_0^t \lambda_{jk}(u)\,du$ denote the basic martingales. The four models to be considered in this paper are described as follows.

MODEL 1 (Cox's model). Assume $\lambda_{j0}(\cdot) = h(\cdot)$ is a deterministic function for every $j = 1, 2, \dots$.

MODEL 2. Assume that

$$\lambda_{j0}(t) = \sum_{i=0}^{\infty} h_i(t) 1_{(T_{ji}, T_{ji+1})]}(t), \tag{2.2}$$

where $h_0, h_1, \dots$ are deterministic functions and $T_{ji} = \inf\{t > 0 \mid \sum_{k=1}^{K} N_{jk}(t) = i\}$. Let $h = (h_o, h_1, \dots)$. Equation (2.2) is slightly more general than (1.3). In this case the shape of the baseline hazard function depends on the number of preceding failures for the study subject. In addition, the baseline intensity $h_i(\cdot)$ for stratum i depends only on t, the time from the beginning of study.

MODEL 3. Assume that

$$\lambda_{j0}(t) = \sum_{i=0}^{\infty} h_i(t - T_{ij}) 1_{(T_{ji}, T_{ji+1})]}(t), \tag{2.3}$$

where $h_0, h_1, \dots$ are nonnegative deterministic functions. Let $h = (h_0, h_1, \dots)$. For stratum $i, h_i(t - T_{ji})$ depends on $t - T_{ji}$, the time from the study subjects' immediately preceding failure.

MODEL 4. Assume that

$$\lambda_{j0}(t) = \sum_{i=0}^{\infty} h_i(T_{j1}, \dots, T_{ji}, X_{j1}, \dots, X_{ji}, t) 1_{(T_{ji}, T_{ji+1}]}(t), \tag{2.4}$$

where $X_{ji} = k$ if $N_{jk}(T_{ji}) - N_{jk}(T_{ji}^-) = 1$ and h_i is a nonnegative deterministic function. Let $h = (h_0, h_1, \dots)$. In this case the shape of the baseline hazard function depends on the whole failure time history, including the number of preceding failures.

It is clear that Model 1 is a special case of Model 2, and both Model 2 and Model 3 are special cases of Model 4. For Model 4, we need to assume $K > 1$.

In all these four models, the statistical problem is to estimate $\theta_0 \in \Theta$, the true parameter value, based on the data

$$\{N_j(t), Y_j(t), Z_j(t) \mid 0 \le t \le t_0,\ 1 \le j \le J\}, \tag{2.5}$$

treating h as a nuisance parameter. Here $Y_j = (Y_{j1}, \dots, Y_{jK}), Z_j = (Z_{j1}, \dots, Z_{jK})$ and t_0 is a finite positive constant, denoting the terminating time.

The following notation helps to describe the nuisance parameter space.

Let

$$Q_i = \{(t_1, \dots, t_i) \mid 0 < t_1 < \cdots < t_i\},$$
$$H_i = \{h_i \mid h_i\colon Q_i \times \{1, \dots, K\}^i \times [0, \infty) \to [a, b] \text{ is a measurable function}\},$$

where $0 < a < b < \infty$.

The nuisance parameter space for Model 4 is $H = H_1 \times H_2 \times \cdots$. We note that H is a complete metric space with metric d defined by

$$\begin{aligned} d(h, g) &= d\big((h_0, h_1, h_2, \dots), (g_0, g_1, \dots)\big) \\ &= \sup_i \|h_i - g_i\|, \end{aligned} \tag{2.6}$$

where $\|h_i - g_i\|$ is the supremum of $|h_i - g_i|$ on its domain.

The nuisance parameter spaces for Models 1, 2 and 3 are regarded as subsets of H and are denoted, respectively, by $H^{(1)}, H^{(2)}$ and $H^{(3)}$; H itself is also denoted by $H^{(4)}$.

3. Information calculations and effective score functions. The concept of asymptotic information in semiparametric models was introduced in Begun, Hall, Huang and Wellner (1983). The main thesis of their paper is that asymptotic lower bounds for estimation of θ with nuisance parameter are determined by the geometry of the scores. Their methods involve (i) the notion of a "Hellinger-differentiable (root-) density" to obtain appropriate scores and (ii) calculation of the "effective score" for θ. Among many interesting ideas and examples, they discussed Cox's regression model for survival data and calculated its information to illustrate their methods. Later, a more powerful method for information calculations was developed and applied again to Cox's model for survival data [Ritov and Wellner (1988)].

In this section we will study, adapting the framework outlined by Begun, Hall, Huang and Wellner (1983), the information concept in generalized proportional hazards models for counting processes, from which the "effective score functions" can be obtained.

To ease the presentation, we begin the discussion with the more general Model 4.

MODEL 4. Let $\mathcal{P}^{(\theta, h)}$ denote the probability measure specified by $\theta \in \Theta$, $h \in H$. Assume that H contains $\mathbf{1}$, which is the element whose ith component is the constant function 1 defined on $Q_i \times \{1, \dots, K\}^i \times [0, \infty)$. Assume $0 \in \Theta$. It follows from the Radon–Nikodym derivative theorem for point processes [cf. Brémaud (1981), pages 166, 187] that

$$\text{(3.1)}\qquad \begin{aligned}\frac{d\mathcal{P}^{(\theta,h)}}{d\mathcal{P}^{(0,1)}} &\equiv L_J(t,\theta,h)\\ &\equiv \prod_{j=1}^{J}\prod_{k=1}^{K} L_{jk}(t,\theta,h),\end{aligned}$$

where L_{jk} satisfies

$$\text{(3.2)}\qquad \begin{aligned}\mathcal{L}_{jk}(t,\theta,h) &\equiv \log L_{jk}(t,\theta,h)\\ &= \int_0^t \log\lambda_{j0}(u)\,dN_{jk}(u) + \int_0^t \theta Z_{jk}(u)\,dN_{jk}(u)\\ &\quad + \int_0^t \Big(1-\lambda_{j0}(u)\exp\big[\theta z_{jk}(u)\big]\Big)Y_{jk}(u)\,du\\ &= \sum_{i=0}^{\infty}\int_0^t \log h_i(u)\,dN_{jki}(u) + \int_0^t \theta Z_{jk}(u)\,dN_{jk}(u)\\ &\quad + \sum_{i=0}^{\infty}\int_0^t \Big(1-h_i(u)\exp\big[\theta Z_{jk}(u)\big]\Big)1_{(T_{ji},T_{ji+1}]}(u)Y_{jk}(u)\,du,\end{aligned}$$

where

$$N_{jki}(u) = 1_{(T_{ji},T_{jk+1}]}(u)N_{jk}(u)$$

and $h_i(u)$ is an abbreviation of $h_i(T_{j1},\dots,T_{ji},X_{j1},\dots,X_{ji},u)$.

We note that (3.1) is the Radon–Nikodym derivative of $\mathcal{P}^{(\theta,h)}$ relative to $\mathcal{P}^{(0,1)}$ on $(\Omega,\mathcal{G}_t^J)$ where $\mathcal{G}_t^J = \otimes_{j=1}^J\mathcal{F}_{j,t}$. Let $\mathcal{M}(\Omega,\mathcal{G}_t^J)$ denote the space of $\mathcal{G}_t^J$ measurable functions on Ω. Let $L_2(\Omega,\mathcal{G}_t^J,\mathcal{P}^{(0,1)})$ be the space of square-integrable elements in $\mathcal{M}(\Omega,\mathcal{G}_t^J)$. It is clear that $L_J^{1/2}(t,\theta,h)\in L_2(\Omega,\mathcal{G}_t^J,\mathcal{P}^{(0,1)})$.

Let $\theta_{(0)},\theta_{(1)},\dots$ in Θ, and let $h_{(0)},h_{(1)},\dots$ in H. Assume that

$$\text{(3.3)}\qquad \begin{aligned}\lim_{J\to\infty}\big|\sqrt{J}\big(\theta_{(J)}-\theta_{(0)}\big)-\delta\big| &= 0,\\ \lim_{J\to\infty}\big\|\sqrt{J}\big(h_{(J)}-h_{(0)}\big)-\beta\big\| &= 0,\end{aligned}$$

where $\delta\in\mathbb{R}$ and $\beta\in\widetilde{H}$. Here $\widetilde{H}$ is defined in the same way as H except that $[a,b]$ is replaced by $\mathbb{R}$ and we require that elements in $\widetilde{H}$ have finite norm $d(h,0)$; $\widetilde{H}^{(1)}$, $\widetilde{H}^{(2)}$ and $\widetilde{H}^{(3)}$ are defined similarly. Sometimes we will use $E(\cdot)$ to denote $E_{(\theta_{(0)},h_{(0)})}(\cdot)$, the expectation taken at the true parameter $(\theta_{(0)},h_{(0)})$, to simplify the notation.

PROPOSITION 3.1. *Let $J_0\geq 1$ be a fixed integer. On $(\Omega,\mathcal{G}_t^{J_0},\mathcal{P}^{(0,1)})$, as J goes to infinity, $\sqrt{J}(L_{J_0}^{1/2}(t,\theta_{(J)},h_{(J)})-L_{J_0}^{1/2}(t,\theta_{(0)},h_{(0)}))$ converges almost surely and in*

L_2 to

$$
\begin{aligned}
&\frac{1}{2}L_{J_0}^{1/2}(t,\theta_{(0)},h_{(0)})\\
(3.4)\qquad &\times\left\{\sum_{j=1}^{J_0}\sum_{k=1}^{K}\int_0^t\left[\delta Z_{jk}(u)+\sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)}1_{(T_{ji},T_{ji+1}]}(u)\right]dM_{jk}(u)\right\},
\end{aligned}
$$

which will be denoted by $\alpha_{J_0}(t,\delta,\beta)$.

The proof for Proposition 3.1 is omitted here, because it is tedious and straightforward. The details can be found in Chang and Hsiung (1991c). We note that Proposition 3.1 holds when $K \geq 1$.

Note that

$$
\begin{aligned}
&\|\alpha_{J_0}(t,\delta,\beta)\|_2^2\\
(3.5)\qquad &=\frac{\delta^2}{4}E_{(\theta_{(0)},h_{(0)})}\left\{\sum_{j=1}^{J_0}\sum_{k=1}^{K}\int_0^t\left[Z_{jk}(u)+\sum_{i=0}^{\infty}\frac{\beta_i(u)}{\delta h_{(0),i}(u)}1_{(T_{ji},T_{ji+1}]}(u)\right]\right.\\
&\qquad\left.\times\, dM_{jk}(u)\right\}^2,
\end{aligned}
$$

where $\|\ \|_2$ denotes the L_2-norm. Let

$$
\sum_{J=1}^{J_0}\sum_{k=1}^{K}\int_0^t\left[Z_{jk}(u)+\sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)}1_{(T_{ji},T_{ji+1}]}(u)\right]dM_{jk}(u)
$$

be denoted by $\tilde{\alpha}_{J_0}(t,\beta)$.

With (3.5) and the corresponding results for other models, we have the following definition.

DEFINITION 3.1.

$$
\begin{aligned}
I_*^{(m)} &\equiv \inf\left\{\frac{4}{\delta^2}\|\alpha_1(t_0,\delta,\beta)\|_2^2 \,\middle|\, \delta\in\mathbb{R},\beta\in\tilde{H}^{(m)}\right\}\\
&\equiv \inf\left\{E_{(\theta_{(0)},h_{(0)})}\big(\tilde{\alpha}_1(t_0,\beta)\big)^2 \,\middle|\, \beta\in\tilde{H}^{(m)}\right\}
\end{aligned}
$$

is called the asymptotic information for estimation of $\theta_{(0)}$ in Model m, $m = 1,2,3,4$.

The quantity $I_*^{(m)}$ is important in the representation theorem for regular estimators. In the following we will use a martingale approach and an appropriate orthogonality argument to calculate $I_*^{(m)}$ for $m = 1,2,3,4$.

Let

$$
(3.6)\qquad S_j^{(q)}(\theta,t)=\sum_{k=1}^{K}Y_{jk}(t)\exp\big[\theta Z_{jk}(t)\big]Z_{jk}^q(t),
$$

where $q = 0, 1, 2$. Consider the estimating function

$$G_J^{(4)}(\theta, t) = \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \left(Z_{jk}(u) - \frac{S_j^{(1)}(\theta, u)}{S_j^{(0)}(\theta, u)}\right) dN_{jk}(u). \tag{3.7}$$

The following theorem establishes a relation between the asymptotic information $I_*^{(4)}$ and $E_{(\theta_{(0)}, h_{(0)})}(G_J^{(4)}(\theta_{(0)}, t_0))^2$.

THEOREM 3.1. *In Model* 4,

$$I_*^{(4)} = \frac{1}{J} E_{(\theta_{(0)}, h_{(0)})}\left(G_J^{(4)}(\theta_{(0)}, t_0)\right)^2, \tag{3.8}$$

for every $J = 1, 2, \dots,$

PROOF. Using the fact that

$$\sum_{k=1}^{K}\int_0^t \left[Z_{jk}(u) - \frac{S_j^{(1)}(\theta_{(0)}, u)}{S_j^{(0)}(\theta_{(0)}, u)}\right] dM_{jk}(u)$$

and

$$\sum_{k=1}^{K}\int_0^t \left[\frac{S_j^{(1)}(\theta_{(0)}, u)}{S_j^{(0)}(\theta_{(0)}, u)} + \sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)} 1_{(T_{ji}, T_{ji+1}]}(u)\right] dM_{jk}(u)$$

are orthogonal martingales [cf. Chang and Hsiung (1991a), (2.17)], where $\beta_i(u)$ and $h_{(0),i},(u)$ are abbreviations of $\beta_i(T_{j1}, \dots, T_{ji}, X_{j1}, \dots, X_{ji}, u)$ and $h_{(0),i}(T_{j1}, \dots, T_{ji}, X_{j1}, \dots, X_{ji}, u)$, respectively, we have, for $\beta \in \widetilde{H}^{(4)}$,

$$\begin{aligned}
(\tilde{\alpha}_1(t_0, \beta))^2 &= \left\{\sum_{k=1}^{K}\int_0^{t_0}\left[Z_{1k}(u) + \sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)} 1_{(T_{1i}, T_{1i+1}]}(u)\right] dM_{1k}(u)\right\}^2 \\
&= \left\{\sum_{k=1}^{K}\int_0^{t_0}\left[Z_{1k}(u) - \frac{S_1^{(1)}(\theta_{(0)}, u)}{S_1^{(0)}(\theta_{(0)}, u)}\right] dM_{1k}(u)\right\}^2 \\
&\quad + \left\{\sum_{k=1}^{K}\int_0^{t_0}\left[\frac{S_1^{(1)}(\theta_{(0)}, u)}{S_1^{(0)}(\theta_{(0)}, u)} + \sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)} 1_{(T_{1i}, T_{1i+1}]}(u)\right] dM_{1k}(u)\right\}^2.
\end{aligned} \tag{3.9}$$

Since predictable processes relative to the self-excited filtration of a counting process admit representations of the form (2.4) [cf. Brémaud (1981), page 309], we know

$$\frac{S_1^{(1)}(\theta_{(0)}, u)}{S_1^{(0)}(\theta_{(0)}, u)} = \sum_{i=0}^{\infty}\frac{\beta_i^*(u)}{h_{(0),i}(u)} 1_{(T_{1i}, T_{1i+1}]}(u),$$

for some $\beta^* = (\beta_0^*, \beta_1^*, \dots) \in \widetilde{H}^{(4)}$. Again $\beta_i^*(u)$ is an abbreviation for $\beta_i^*(T_{j1}, \dots, T_{ji}, X_{j1}, \dots, X_{ji}, u)$. This together with (3.9) shows that the minimum of

$E_{(\theta_{(0)}, h_{(0)})}(\widetilde{\alpha}_1(t_0, \beta))^2$ is attained by choosing $\beta_i(u) = -\beta_i^*(u)$, $i = 0, 1, 2, \ldots$. Hence $I_*^{(4)}$ is equal to the first term in (3.9), which is in fact $E_{(\theta_{(0)}, h_{(0)})}(G_1^{(4)}(\theta_{(0)}, t_0))^2$. A straightforward calculation using again the orthogonality of certain martingales shows that (3.8) holds for $J \geq 1$. This completes the proof. □

The proof of Theorem 3.1 indicates that the "effective score" in Model 4 is precisely $G_J^{(4)}(\theta, t)$, the estimating function proposed in Chang and Hsiung (1991a). This is mainly due to the fact that the enlargement of the nuisance parameter space entails the existence of $\beta^* \in \widetilde{H}^{(4)}$ for the representation of $S_1^{(1)}(\theta_{(0)}, u)/S_1^{(0)}(\theta_{(0)}, u)$. We note also that if $K = 1$, then $G_J^{(4)}(\theta, t) \equiv 0$. This is the reason we need $K > 1$ in Model 4.

MODEL 1 (Cox's model). Let

$$S_J^{(q)}(\theta, t) = \frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{jk}(t) \exp\left[\theta Z_{jk}(t)\right] Z_{jk}^q(t) \tag{3.10}$$

where $q = 0, 1, 2$, and let

$$s^{(q)}(\theta, t) = E \sum_{k=1}^{K} Y_{1k}(t) \exp\left[\theta Z_{1k}(t)\right] Z_{1k}^q(t). \tag{3.11}$$

With (3.10), we can express the Cox MPL equation as

$$G_J^{(1)}(\theta, t) = \sum_{j=1}^{J} \sum_{k=1}^{K} \int_0^t \left(Z_{jk}(u) - \frac{S_J^{(1)}(\theta, u)}{S_J^{(0)}(\theta, u)} \right) dN_{jk}(u). \tag{3.12}$$

The following theorem gives the relation between the asymptotic information $I_*^{(1)}$ and $E_{(\theta_{(0)}, h_{(0)})}(G_J^{(1)}(\theta_{(0)}, t_0))^2$.

THEOREM 3.2. *In Cox's model,*

$$I_*^{(1)} = E_{(\theta_{(0)}, h_{(0)})} \left(\sum_{k=1}^{K} \int_0^{t_0} \left(Z_{1k}(u) - \frac{s^{(1)}(\theta_{(0)}, u)}{s^{(0)}(\theta_{(0)}, u)} \right) dM_{1k}(u) \right)^2, \tag{3.13}$$

$$= \lim_{J \to \infty} \frac{1}{J} E_{(\theta_{(0)}, h_{(0)})} \Big(G_J^{(1)}(\theta_{(0)}, t_0) \Big)^2. \tag{3.14}$$

PROOF. Let $\beta \in \widetilde{H}^{(1)}$ and $h_{(0)}) \in H^{(1)}$. Using the fact that

$$\sum_{j,k} \int_0^t \left(Z_{jk}(u) - \frac{S_J^{(1)}(\theta_{(0)}, u)}{S_J^{(0)}(\theta_{(0)}, u)} \right) dM_{jk}$$

and

$$\sum_{j,k}\int_0^t \left(\frac{S_J^{(1)}(\theta_{(0)},u)}{S_J^{(0)}(\theta_{(0)},u)} + \frac{\beta(u)}{h_{(0)}(u)}\right) dM_{jk}$$

are orthogonal martingales, we have

$$\begin{aligned} &E\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(u)+\frac{\beta(u)}{h_{(0)}(u)}\right)dM_{1k}(u)\right)^2 \\ (3.15)\qquad &\geq E\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(u)-\frac{S_J^{(1)}(\theta_{(0)},u)}{S_J^{(0)}(\theta_{(0)},u)}\right)dM_{1k}(u)\right)^2 \\ &= \frac{1}{J}E\left(G_J^{(1)}(\theta_{(0)},t_0)\right)^2. \end{aligned}$$

Taking limit in each of the three expressions in (3.15), we get

$$(3.16)\quad \inf_{\beta\in\tilde{H}^{(1)}} E\left(\tilde{\alpha}_1(t_0,\beta)\right)^2 = E\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(u)-\frac{s^{(1)}(\theta_{(0)},u)}{s^{(0)}(\theta_{(0)},u)}\right)dM_{1k}(u)\right)^2$$

This completes the proof. □

Next we have similar results for Model 2, the proof of which can be found in Chang and Hsiung (1991c).

MODEL 2. Let

$$(3.17)\qquad S_{i,J}^{(q)}(\theta,t) = \frac{1}{J}\sum_{j=1}^{J}\sum_{k=1}^{K} 1_{(T_{ji},T_{ji+1}]}(t)Y_{jk}(t)e^{\theta Z_{jk}(t)}Z_{jk}^{q}(t),$$

where $q = 0, 1, 2$. Let

$$(3.18)\qquad s_i^{(q)}(\theta,t) = ES_{i,1}^{(q)}(\theta,t).$$

Consider the estimating function

$$(3.19)\quad G_J^{(2)}(\theta,t) = \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t\left(Z_{jk}(u)-\sum_{i=0}^{\infty}\frac{S_{i,J}^{(1)}(\theta,u)}{S_{i,J}^{(0)}(\theta,u)}1_{(T_{ji},T_{ji+1}]}(u)\right)dN_{jk}(u).$$

Then we have the following theorem.

THEOREM 3.3. *In Model* 2,

$$(3.20)\quad I_*^{(2)} = E_{(\theta_{(0)},h_{(0)})}\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(u)-\sum_{i=0}^{\infty}\frac{s_i^{(1)}(\theta_{(0)},u)}{s_i^{(0)}(\theta_{(0)},u)}1_{(T_{1i},T_{1i+1}]}(u)\right)dM_{1k}(u)\right)^2$$

$$(3.21)\quad = \lim_{J\to\infty}\frac{1}{J}E_{(\theta_{(0)},h_{(0)})}\left(G_J^{(2)}(\theta_{(0)},t_0)\right)^2.$$

One consequence of Theorems 3.2 and 3.3 is that in Models 1 and 2 the "effective scores" are

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t\left(Z_{jk}(u)-\frac{s^{(1)}(\theta,u)}{s^{(0)}(\theta,u)}\right)dM_{jk}(u)$$

and

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t\left(Z_{jk}(u)-\sum_{i=0}^{\infty}\frac{s_i^{(1)}(\theta,u)}{s_i^{(0)}(\theta,u)}1_{(T_{ji},T_{ji+1})}(u)\right)dM_{jk}(u),$$

respectively, which suggests the use of $G_J^{(1)}(\theta,t)$ and $G_J^{(2)}(\theta,t)$ as estimating functions.

MODEL 3. The relevant martingale structures in Model 3 are different from those in Models 1, 2 and 4. We will exhibit some martingales relative to filtrations other than the self-excited one. By examining carefully the definition of asymptotic information given in Definition 3.1, we are led to explore the following filtrations and martingales.

Let

$$M_{jk}(t)=N_{jk}(t\wedge t_0)-\int_0^t\lambda_{j0}(s)1_{(0,t_0]}(s)Y_{jk}(s)\exp\left[\theta Z_{jk}(s)\right]ds. \tag{3.22}$$

Since $M_{jk}(t)$ is an $\mathcal{F}_{j,t}$-martingale, we know $M_{jk}(T_{ji}+t)$ is an $\mathcal{F}_{j,T_{ji}+t}$-martingale. This together with the fact that $T_{ji+1}-T_{ji}$ is an $\mathcal{F}_{j,T_{ji}+t}$-stopping time shows that

$$M_{jki}(t)\equiv M_{jk}\left(T_{ji}+t\wedge(T_{ji+1}-T_{ji})\right)-M_{jk}(T_{ji}) \tag{3.23}$$

is also an $\mathcal{F}_{j,T_{ji}+t}$-martingale.

Let

$$N_{jki}(t)=N_{jk}\Big(\big(T_{ji}+t\wedge(T_{ji+1}-T_{ji})\big)\wedge t_0\Big)-N_{jk}(T_{ji}\wedge t_0).$$

Observe that

$$\begin{aligned}M_{jki}(t)&=N_{jki}(t)-\int_{T_{ji}}^{T_{ji}+t\wedge(T_{ji+1}-T_{ji})}\lambda_{j0}(s)1_{(0,t_0]}(s)Y_{jk}(s)\exp\left[\theta Z_{jk}(s)\right]ds\\&=N_{jki}(t)-\int_{T_{ji}}^{T_{ji}+t\wedge(T_{ji+1}-T_{ji})}h_i(s-T_{ji})1_{(0,t_0]}(s)Y_{jk}(s)\exp\left[\theta Z_{jk}(s)\right]ds\\&=N_{jki}(t)-\int_0^t h_i(u)1_{(0,T_{ji+1}-T_{ji}]}(u)1_{(0,t_0]}(T_{ji}+u)\\&\qquad\times Y_{jk}(T_{ji}+u)\exp\left[\theta Z_{jk}(T_{ji}+u)\right]du.\end{aligned} \tag{3.24}$$

Let

$$\mu_{jki}(\theta,u)=h_i(u)1_{(0,T_{ji+1}-T_{ji}]}(u)1_{(0,t_0]}(T_{ji}+u)\times Y_{jk}(T_{ji}+u)\exp\left[\theta Z_{jk}(T_{ji}+u)\right].$$

Since the integrand in (3.24) is a predictable process relative to $\mathcal{F}_{j,T_{ji}+t}$, we know that, for each i, $\{N_{jki}(t); j=1,\dots,J, k=1,\dots,K\}$ is a JK-variate counting process with proportional hazards relative to the filtration

$$\mathcal{F}^i_{(J),t}\equiv\sigma\{\mathcal{F}_{j,T_{ji}+t}\mid j=1,\dots,J\},$$

and the common baseline function is h_i. This is a Cox proportional hazard model. Therefore, if we were to use only information contained in $\mathcal{F}^i_{(J),t_0}$, we would certainly use the estimating function

$$G^i_J(\theta,t)=\sum_{j=1}^J\sum_{k=1}^K\int_0^t\left(Z_{jk}(T_{ji}+u)-\frac{\widetilde{S}^{(1)}_{iJ}(\theta,u)}{\widetilde{S}^{(0)}_{iJ}(\theta,u)}\right)dN_{jki}(u),\tag{3.25}$$

where

$$\widetilde{S}^{(q)}_{iJ}(\theta,u)=\frac{1}{J}\sum_{j=1}^J\sum_{k=1}^K 1_{(0,T_{ji+1}-T_{ji}]}(u)1_{(0,t_0]}(T_{ji}+u)\times Y_{jk}(T_{ji}+u)\exp\left[\theta Z_{jk}(T_{ji}+u)\right]Z^q_{jk}(T_{ji}+u),$$

with $q=0,1,2$. Let

$$\widetilde{s}^{(q)}_i(\theta,u)=E\widetilde{S}^{(q)}_{i1}(\theta,u),$$

for $q=0,1,2$.

Since for $i\neq j$, $G^i_J(\theta_{(0)},t_0)$ and $G^j_J(\theta_{(0)},t_0)$ are uncorrelated (cf. Lemma 3.1), we are encouraged to combine these estimating functions and define

$$G^{(3)}_J(\theta,t_0)\equiv\sum_{i=0}^{\infty}G^i_J(\theta,t_0),\tag{3.26}$$

which hopefully may provide approximation to the asymptotic information $I^{(3)}_*$. In fact, (3.26) was also proposed by Prentice, Williams and Petersen (1981) in estimation of θ and they expressed interest in developing asymptotic estimation theory based on (3.26).

LEMMA 3.1. *For each fixed J, $\{\sum_{i=0}^I G^i_J(\theta_{(0)},t_0)\mid I=0,1,\dots\}$ is a martingale relative to the filtration $\mathcal{F}^{I+1}_{(J),0}$.*

PROOF. It is obvious that $G^i_J(\theta_{(0)},t_0)$ is $\mathcal{F}^{i+1}_{(J),0}$-measurable. Besides, using the fact that $G^i_J(\theta_{(0)},t)$ is an $\mathcal{F}^i_{(J),t}$-martingale [cf. (3.25)], we know

$$E\left(G^i_J(\theta_{(0)},t_0)\mid\mathcal{F}^i_{(J),0}\right)=G^i_J\left(\theta_{(0)},0\right)=0.$$

This completes the proof. □

It follows from Lemma 3.1 and the quadratic variation formula for martingales that

$$E\left(\sum_{i=0}^{I} G_J^i(\theta_{(0)}, t_0)\right)^2 = \sum_{i=0}^{I} E\big(G_J^i(\theta_{(0)}, t_0)\big)^2 \tag{3.27}$$
$$\le \sum_{i=0}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} C \cdot E(T_{ji+1} - T_{ji}) \wedge t_0$$
$$\le C \cdot JK \cdot t_0,$$

for some constant C, for every I.

THEOREM 3.4. *In Model* 3,

$$I_*^{(3)} = \sum_{i=0}^{\infty} E\left(\sum_{k=1}^{K} \int_0^{t_0} \left(Z_{1k}(T_{1i} + u) - \frac{\tilde{s}_i^{(1)}(\theta_{(0)}, u)}{\tilde{s}_i^{(0)}(\theta_{(0)}, u)}\right) dM_{1ki}(u)\right)^2 \tag{3.28}$$

$$= \lim_{J \to \infty} \frac{1}{J} E\big(G_J^{(3)}(\theta_{(0)}, t_0)\big)^2. \tag{3.29}$$

PROOF. Using Lemma 3.1, (3.27) and the fact that the martingales

$$\sum_{j=1}^{J}\sum_{k=1}^{K} \int_0^t \left(Z_{jk}(T_{ji} + u) - \frac{\tilde{S}_{iJ}^{(1)}(\theta_{(0)}, u)}{\tilde{S}_{iJ}^{(0)}(\theta_{(0)}, u)}\right) dM_{jki}(u)$$

and

$$\sum_{j=1}^{J}\sum_{k=1}^{K} \int_0^t \left(\frac{\tilde{S}_{iJ}^{(1)}(\theta_{(0)}, u)}{\tilde{S}_{iJ}^{(0)}(\theta_{(0)}, u)} - \frac{\beta_i(u)}{h_{(0),i}(u)}\right) dM_{jki}(u)$$

are orthogonal, we have

$$\frac{1}{J} E\left(G_J^{(3)}(\theta_{(0)}, t_0)\right)^2$$
$$= \frac{1}{J} \sum_{i=0}^{\infty} E\left(G_J^i(\theta_{(0)}, t_0)\right)^2$$
$$= \frac{1}{J} \sum_{i=0}^{\infty} E\left(\sum_{j=1}^{J}\sum_{k=1}^{K} \int_0^{t_0} \left(Z_{jk}(T_{ji} + u) - \frac{\tilde{S}_{iJ}^{(1)}(\theta_{(0)}, u)}{\tilde{S}_{iJ}^{(0)}(\theta_{(0)}, u)}\right) dM_{jki}(u)\right)^2 \tag{3.30}$$
$$= \sum_{i=0}^{\infty} E\left(\sum_{k=1}^{K} \int_0^{t_0} \left(Z_{1k}(T_{1i} + u) - \frac{\tilde{S}_{iJ}^{(1)}(\theta_{(0)}, u)}{\tilde{S}_{iJ}^{(0)}(\theta_{(0)}, u)}\right) dM_{1ki}(u)\right)^2$$
$$< \frac{1}{J} \sum_{i=0}^{\infty} E\left(\sum_{j=1}^{J}\sum_{k=1}^{K} \int_0^{t_0} \left(Z_{jk}(T_{ji} + u) - \frac{\beta_i(u)}{h_{(0),i}(u)}\right) dM_{jki}(u)\right)^2.$$

Taking the limit in (3.30), we get

$$\lim_{J\to\infty}\frac{1}{J}E\Big(G_J^{(3)}(\theta_{(0)},t_0)\Big)^2$$

$$(3.31)\qquad = \sum_{i=0}^{\infty}E\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(T_{1i}+u)-\frac{\tilde{s}_i^{(1)}(\theta_{(0)},u)}{\tilde{s}_i^{(0)}(\theta_{(0)},u)}\right)dM_{1ki}(u)\right)^2$$

$$\leq \sum_{i=0}^{\infty}E\left(\sum_{k=1}^{K}\int_0^{t_0}\left(Z_{1k}(T_{1i}+u)-\frac{\beta_i(u)}{h_{(0),i}(u)}\right)dM_{1ki}(u)\right)^2.$$

The proof is complete. □

4. Asymptotic normality. Let $\widehat{\theta}_J^{(m)}$ be the solution of $G_J^{(m)}(\cdot,t_0)=0$; $\widehat{\theta}_J^{(m)}$ is called the maximum partial likelihood estimator (MPLE) of the relative risk coefficient for Model m. The purpose of this section is to establish the asymptotic normality of $\widehat{\theta}_J^{(m)}$ in Model m.

There are two subsections in this section. Section 4.1, treats Models 1, 2 and 4; Section 4.2 treats Model 3. Since the asymptotic normality of $\widehat{\theta}_J^{(1)}$ was established by Andersen and Gill (1982) and the asymptotic normality of $\widehat{\theta}_J^{(m)}$, for $m = 2, 4$, can be established in a similar, although more complicated manner, we will omit the proofs and give only the statements with important formulas for later use. The detailed proofs can be found in Chang and Hsiung (1991c). The case for $\widehat{\theta}_J^{(3)}$ involves deeper analysis of the martingale structures. We will supply some details for it.

4.1. *Asymptotic normality in Models* 1, 2 *and* 4.

MODEL 1 (Cox's model). Andersen and Gill (1982) discussed this case in a more general setting. We will present it in a form suitable for variance comparison.

Observe that the quadratic variation process

$$(4.1)\qquad \left\langle\frac{1}{\sqrt{J}}G_J^{(1)}(\theta_0,\cdot)\right\rangle_t=\int_0^t h(u)\left(S_J^{(2)}(\theta_0,u)-\frac{\big(S_J^{(1)}(\theta_0,u)\big)^2}{S_J^{(0)}(\theta_0,u)}\right)du,$$

which converges in probability to

$$(4.2)\qquad V_1(t)=\int_0^t h(u)\left(s^{(2)}(\theta_0,u)-\frac{\big(s^{(1)}(\theta_0,u)\big)^2}{s^{(0)}(\theta_0,u)}\right)du,$$

where $S_J^{(q)}(\theta,t)$ and $s^{(q)}(\theta,t)$ are defined in (3.10) and (3.11), respectively.

PROPOSITION 4.1. *The process* $(1/\sqrt{J})G_J^{(1)}(\theta_0,t)$ *converges weakly on* $D[0,t_0]$ *to a continuous Gaussian martingale with variation process* $V_1(t)$ *given in* (4.2).

THEOREM 4.1. *If* $V_1(t_0) > 0$, *then* $\sqrt{J}(\widehat{\theta}_J^{(1)} - \theta_0)$ *is asymptotically normal with mean* 0 *and variance* $(V_1(t_0))^{-1}$.

MODEL 2. Let $S_{i,J}^{(q)}(\theta,t)$ and $s_i^{(q)}(\theta,t)$ be defined as in (3.17) and (3.18), respectively. Then on $\Theta \times [0,t_0]$, $S_{i,J}^{(q)}(\theta,t)$ converges to $s_i^{(q)}(\theta,t)$ uniformly with probability 1.

When $\theta = \theta_0$, we know

$$G_J^{(2)}(\theta_0,t) = \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \left(Z_{jk}(u) - \sum_{i=0}^{\infty}\frac{S_{iJ}^{(1)}(\theta_0,u)}{S_{iJ}^{(0)}(\theta_0,u)}1_{(T_{ji},T_{ji+1}]}(u)\right)dM_{jk}(u), \tag{4.3}$$

and the quadratic variation process

$$\left\langle \frac{1}{\sqrt{J}}G_J^{(2)}(\theta_0,\cdot)\right\rangle_t = \sum_{i=0}^{\infty}\int_0^t h_i(u)\left(S_{iJ}^{(2)}(\theta_0,u) - \frac{\left(S_{iJ}^{(1)}(\theta_0,u)\right)^2}{S_{iJ}^{(0)}(\theta_0,u)}\right)du, \tag{4.4}$$

which converges in probability to

$$V_2(t) = \sum_{i=0}^{\infty}\int_0^t h_i(u)\left(s_i^{(2)}(\theta_0,u) - \frac{\left(s_i^{(1)}(\theta_0,u)\right)^2}{s_i^{(0)}(\theta_0,u)}\right)du. \tag{4.5}$$

PROPOSITION 4.2. *The process* $(1/\sqrt{J})G_J^{(2)}(\theta_0,t)$ *converges weakly on* $D[0,t_0]$ *to a continuous Gaussian martingale with variation process* $V_2(t)$ *given in* (4.5).

THEOREM 4.2. *If* $V_2(t_0) > 0$, *then* $\sqrt{J}(\widehat{\theta}_J^{(2)} - \theta_0)$ *is asymptotically normal with mean* 0 *and variance* $(V_2(t_0))^{-1}$.

MODEL 4. Let $S_j^{(q)}(\theta,t)$ be defined as in (3.6). We know that the quadratic variation process

$$\left\langle \frac{1}{\sqrt{J}}G_J^{(4)}(\theta_0,\cdot)\right\rangle_t = \frac{1}{J}\sum_{j=0}^{J}\int_0^t \lambda_{j0}(u)\left(S_j^{(2)}(\theta_0,u) - \frac{\left(S_j^{(1)}(\theta_0,u)\right)^2}{S_j^{(0)}(\theta_0,u)}\right)du, \tag{4.6}$$

which converges in probability to

$$V_4(t) = \int_0^t E\lambda_{10}(u)\left(S_1^{(2)}(\theta_0,u) - \frac{\left(S_1^{(1)}(\theta_0,u)\right)^2}{S_1^{(0)}(\theta_0,u)}\right)du. \tag{4.7}$$

PROPOSITION 4.3. *The process* $(1/\sqrt{J})G_J^{(4)}(\theta_0,t)$ *converges weakly to a continuous Gaussian martingale with variation process* $V_4(t)$ *given in* (4.7).

THEOREM 4.3. *If* $V_4(t_0) > 0$, *then* $\sqrt{J}(\widehat{\theta}_J^{(4)} - \theta_0)$ *is asymptotically normal with mean* 0 *and variance* $(V_4(t_0))^{-1}$.

4.2. *Asymptotic consistency and normality for Model* 3.

4.2.1. *Asymptotic consistency.* Let

$$C_J^i(\theta,t) = \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \Big(\theta Z_{jk}(T_{ji}+u) - \log\big(J\widetilde{S}_{iJ}^{(0)}(\theta,u)\big)\Big)dN_{jki}(u),$$

$$X_J^i(\theta,t) = \frac{1}{J}\big(C_J^i(\theta,t) - C_J^i(\theta_{(0)},t)\big)$$
$$= \frac{1}{J}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \left((\theta-\theta_{(0)})Z_{jk}(T_{ji}+u) - \log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\right)dN_{jki}(u),$$

where $\widetilde{S}_{iJ}^{(0)}$ is given in (3.25).

To get the asymptotic consistency of $\widehat{\theta}_J^{(3)}$, it suffices to show that the maximum of $\sum_{i=0}^{\infty} X_J^i(\cdot,t_0)$ converges to $\theta_{(0)}$ in probability. For this, we will first establish three lemmas.

Let

$$A_J^i(\theta,t) = \frac{1}{J}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \left((\theta-\theta_{(0)})Z_{jk}(T_{ji}+u) - \log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\right)\mu_{jki}(\theta_{(0)},u)\,du$$
$$= \int_0^t \left((\theta-\theta_{(0)})\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u) - \log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)\right)h_i(u)\,du,$$

where $\log(0/0) \equiv 0$.

It is easily seen that, for $t \geq 0, X_J^i(\theta,t) - A_J^i(\theta,t)$ is an $\mathcal{F}_{(J),t}^i$-martingale. As a consequence, we can calculate its quadratic variation process as follows:

$$B_J^i(\theta,t) \equiv \langle X_J^i(\theta,\cdot) - A_J^i(\theta,\cdot)\rangle_t$$
$$= \frac{1}{J^2}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \left((\theta-\theta_{(0)})Z_{jk}(T_{ji}+u) - \log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\right)^2 \times \mu_{jki}(\theta_{(0)},u)\,du, \tag{4.8}$$

which converges to 0 in probability as J goes to infinity.

Another important consequence of the fact that $X_J^i(\theta,t) - A_J^i(\theta,t)$ is an $\mathcal{F}_{(J),t}^i$-martingale is the following martingale structure.

LEMMA 4.1. *For fixed* J, $\{\sum_{i=1}^{I}(X_J^i(\theta,t_0) - A_J^i(\theta,t_0)) \mid I = 0,1,\ldots\}$ *is a martingale relative to the filtration* $\mathcal{F}_{(J),0}^{I+1}$.

We will omit the proof for Lemma 4.1 because it can be shown in the same manner as Lemma 3.1.

It follows from Lemma 4.1 and (4.8) that

$$E\left(\sum_{i=0}^{I}\left(X_J^i(\theta,t_0)-A_J^i(\theta,t_0)\right)\right)^2$$

$$=\sum_{i=0}^{I}E\left(X_J^i(\theta,t_0)-A_J^i(\theta,t_0)\right)^2$$

$$(4.9)\qquad =\frac{1}{J^2}\sum_{i=0}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}E\int_0^{t_0}\left((\theta-\theta_{(0)})Z_{jk}(T_{ji}+u)-\log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\right)^2 \times\mu_{jki}(\theta_{(0)},u)\,du$$

$$\le\frac{1}{J^2}\cdot J\cdot C\cdot E\sum_{i=0}^{I}\left(T_{ji+1}-T_{ji}\right)\wedge t_0$$

$$\le\frac{1}{J}\cdot C\cdot t_0,$$

for some constant C, for every I.

Relation (4.9) implies that

$$(4.10)\qquad E\left(\sum_{i=0}^{\infty}\left(X_J^i(\theta,t_0)-A_J^i(\theta,t_0)\right)\right)^2\le\frac{1}{J}\cdot C\cdot t_0.$$

This shows that the following lemma holds.

LEMMA 4.2. *As J goes to infinity, $\sum_{i=0}^{\infty}(X_J^i(\theta,t_0)-A_J^i(\theta,t_0))$ converges to 0 in probability.*

Since

$$\sum_{i=0}^{\infty}\left|A_J^i(\theta,t)\right|$$

$$(4.11)\qquad \le\frac{1}{J}\sum_{j=1}^{J}\sum_{i=0}^{\infty}\sum_{k=1}^{K}\int_0^{t_0}\left|(\theta-\theta_{(0)})Z_{jk}(T_{ji}+u)-\log\frac{\widetilde{S}_{iJ}^{(0)}(\theta,u)}{\widetilde{S}_{iJ}^{(0)}(\theta_{(0)},u)}\right|\mu_{jki}(\theta_{(0)},u)\,du$$

$$\le\frac{1}{J}\sum_{j=1}^{J}C\sum_{i=0}^{\infty}(T_{ji+1}-T_{ji})\wedge t_0$$

$$\le C\cdot t_0,$$

we know from the law of large numbers that the following lemma holds.

LEMMA 4.3. *As J goes to infinity, $\sum_{i=0}^{\infty} A_J^i(\theta, t_0)$ converges to $A(\theta, t_0)$ in probability, where $A(\theta, t_0) = \sum_{i=0}^{\infty} A^i(\theta, t_0)$,*

$$A^i(\theta, t_0) = \int_0^{t_0} \left((\theta - \theta_{(0)})\tilde{s}_i^{(1)}(\theta_{(0)}, u) - \log \frac{\tilde{s}_i^{(0)}(\theta, u)}{\tilde{s}_i^{(0)}(\theta_{(0)}, u)} \tilde{s}_i^{(0)}(\theta_{(0)}, u) \right) h_i(u)\, du$$

and $\tilde{s}_i^{(q)}(\theta, u) = E\tilde{S}_{i1}^{(q)}(\theta, u)$.

It follows from Lemma 4.2, Lemma 4.3, the concaveness of $\sum_{i=0}^{\infty} X_J^i(\cdot, t_0)$ and of $A(\cdot, t_0)$ and the convex analysis argument used in Andersen and Gill (1982) that we have the following consistency result. In fact, $A(\cdot, t_0)$ is a concave function and has a unique maximum at $\theta = \theta_{(0)}$ if $V_3(t_0) > 0$; $\hat{\theta}_J^{(3)}$ maximizes the random concave function $\sum_{i=0}^{\infty} X_J^i(\cdot, t_0)$, which converges to $A(\cdot, t_0)$ in probability; therefore we have Theorem 4.4.

THEOREM 4.4. *The MPLE $\hat{\theta}_J^{(3)}$ converges to $\theta_{(0)}$ in probability if $V_3(t_0) > 0$, where*

$$V_3(t_0) = \sum_{i=0}^{\infty} \int_0^{t_0} h_i(u) \left(\tilde{s}_i^{(2)}(\theta_{(0)}, u) - \frac{\left(\tilde{s}_i^{(1)}(\theta_{(0)}, u)\right)^2}{\tilde{s}_i^{(0)}(\theta_{(0)}, u)} \right) du. \tag{4.12}$$

4.2.2. *Asymptotic normality.* Asymptotic normality of $\hat{\theta}_J^{(3)}$ is to be established in the usual manner by making use of the asymptotic consistency of $\hat{\theta}_J^{(3)}$ and the asymptotic normality of $(1/\sqrt{J})G_J^{(3)}(\theta_{(0)}, t_0)$, which requires some additional martingale structres and tightness argument, similar to those used in Section 4.2.1.

THEOREM 4.5. *The process $(1/\sqrt{J})G_J^{(3)}(\theta_{(0)}, t_0)$ is asymptotically normal with mean 0 and variance $V_3(t_0)$ given in (4.12).*

PROOF. Since $G_J^i(\theta_{(0)}, \cdot)$ is a martingale, we can use martingale central limit theorem to show that $(1/\sqrt{J})G_J^i(\theta_{(0)}, t_0)$ is asymptotically normal with mean 0 and variance

$$V^i(t_0) = \int_0^{t_0} h_i(u) \left(\tilde{s}_i^{(2)}(\theta_{(0)}, u) - \frac{\left(\tilde{s}_i^{(1)}(\theta_{(0)}, u)\right)^2}{\tilde{s}_i^{(0)}(\theta_{(0)}, u)} \right) du. \tag{4.13}$$

On the other hand, we know from (3.27) that, for every I,

$$\left\{ \frac{1}{\sqrt{J}} \left(G_J^0(\theta_{(0)}, t_0), \ldots, G_J^I(\theta_{(0)}, t_0) \right) \middle| \, J = 1, 2, \ldots \right\}$$

is a tight sequence of $\mathbb{R}^{I+1}$-valued random vectors. This together with the previous asymptotical normality of $(1/\sqrt{J})G_J^i(\theta_{(0)}, t_0)$ and the orthogonality

$$EG_J^i(\theta_{(0)}, t_0)G_J^j(\theta_{(0)}, t_0) = 0$$

for $i \neq j$, shows that

$$\frac{1}{\sqrt{J}}\Big(G_J^0(\theta_{(0)}, t_0), \dots, G_J^I(\theta_{(0)}, t_0)\Big)$$

is asymptotically normal with mean 0 and a diagonal covariance matrix whose (i,i)th entry is (4.13).

Therefore, $(1/\sqrt{J})\sum_{i=0}^I G_J^i(\theta_{(0)}, t_0)$ is asymptotically normal with mean 0 and variance $\sum_{i=0}^I V^i(t_0)$.

Consider

$$\begin{aligned} &P\left[\frac{1}{\sqrt{J}}G_J^{(3)}(\theta_{(0)}, t_0) < x\right] \\ &\quad \le P\left[\frac{1}{\sqrt{J}}\sum_{i=0}^{I} G_J^i(\theta_{(0)}, t_0) < x + \varepsilon\right] + P\left[\left|\frac{1}{\sqrt{J}}\sum_{i=I+1}^{\infty} G_J^i(\theta_{(0)}, t_0)\right| \ge \varepsilon\right], \end{aligned} \tag{4.14}$$

where the second term can be made small by (3.27) and the choice of a large I, and this large I will make the first term close to the desired quantity. This shows that (4.14) has its limit supremum equal to $P(X < x)$, where X is normal with mean 0 and variance $V_3(t_0)$. A similar argument gives the corresponding result for the limit infimum of (4.14). This completes the proof. □

With Theorems 4.4 and 4.5, we get the following theorem.

THEOREM 4.6. *$\sqrt{J}(\widehat{\theta}_J^{(3)} - \theta_{(0)})$ is asymptotically normal with mean 0 and variance $V_3(t_0)^{-1}$.*

5. Asymptotic efficiency. In this section, we will show that the MPLE $\widehat{\theta}_J^{(m)}$ is asymptotically efficient in Model m, for $m = 1, 2, 3, 4$. We will establish the convolution theorems in these models and prove that $\widehat{\theta}_J^{(m)}$ are regular estimators, which together with the information calculations in Section 3 gives the asymptotic efficiency of $\widehat{\theta}_J^{(m)}$.

The proof of the convolution theorems follows the well-known characteristic function approach, which requires the local asymptotic normality (LAN) of log-likelihood ratios [cf. Begun, Hall, Huang and Wellner (1983)]. We will present LAN and the convolution theorems without giving the proof. We refer the readers to Chang and Hsiung (1991c) for details. Since the regularity of $\widehat{\theta}_J^{(m)}$ requires a more refined form of LAN, we will present a detailed proof for it in the case of Model 2.

With necessary changes for $h_{(0)}$ and β, the following theorem for LAN holds for all the models. Let

$$(5.1)\qquad \sigma^2(t) = E\sum_{k=1}^{K}\int_0^t \left(\delta Z_{1k}(u) + \sum_{i=0}^{\infty}\frac{\beta_i(u)}{h_{(0),i}(u)}1_{(T_{1i},T_{1i+1}]}(u)\right)^2 \lambda_{1k}(u)\,du.$$

Let $\mathcal{L}_J = \log L_J$. Then we have the following.

THEOREM 5.1 (LAN). *Under $P^{(\theta_{(0)},h_{(0)})}$, as J goes to infinity,*

$$\mathcal{L}_J\big(t,\theta_{(J)},h_{(J)}\big) - \mathcal{L}_J\big(t,\theta_{(0)},h_{(0)}\big) + \tfrac{1}{2}\sigma^2(t)$$

converges weakly on $D(0,t_0]$ to a continuous Gaussian martingale with mean 0 and variance process $\sigma^2(t)$, where $\theta_{(J)}$ and $h_{(J)}$ satisfy (3.3).

Let $\widetilde{\theta}_J$ be an estimator for $\theta_{(0)}$ based on (2.5). We say $\widetilde{\theta}_J$ is a regular estimator at $(\theta_{(0)},h_{(0)})$ if, for every sequence $\{(\theta_{(J)},h_{(J)}\}$ of parameters satisfying (3.3), the distribution of $\sqrt{J}(\widetilde{\theta}_J-\theta_{(J)})$, under $(\theta_{(J)},h_{(J)})$, converges weakly to a distribution which depends on $(\theta_{(0)},h_{(0)})$, but not on the particular sequence $\{(\theta_{(J)},h_{(J)}\}$.

THEOREM 5.2. *Let $\widetilde{\theta}_J$ be a regular estimator of $\theta_{(0)}$ in the model m, $m = 1,2,3,4$. Then, the distribution of $\sqrt{J}(\widetilde{\theta}_J-\theta_{(J)})$, under $(\theta_{(J)},h_{(J)})$, converges weakly to a distribution which is the convolution of $N(0,(I_*^{(m)})^{-1})$ with a distribution W, where W depends only on $(\theta_{(0)},h_{(0)})$ and $I_*^{(m)}$ is given in Definition* 3.1.

THEOREM 5.3. *The solution $\widehat{\theta}_J^{(m)}$ of $G_J^{(m)}(\cdot,t_0) = 0$ is a regular estimator of $\theta_{(0)}$ in Model m, $m = 1,2,3,4$.*

Because of $I_*^{(m)} = V_m(t_0)$, for $m = 1,2,3,4$, and Theorems 5.2 and 5.3, we have the following corollary.

COROLLARY 5.1. *The estimator $\widehat{\theta}_J^{(m)}$ is asymptotically efficient in Model m in the sense that it has the least asymptotic variance among regular estimators, where $m = 1,2,3,4$.*

We note that, when $m = 1$, Theorem 5.2 was established by Begun, Hall, Huang and Wellner (1983) for lifetime data. The fact that $\widehat{\theta}_J^{(1)}$ is regular seems not to be mentioned there. We will prove only the case $m = 2$, for the other cases are similar.

PROOF OF THEOREM 5.3 (Model 2). Let $(\theta_{(J)},h_{(J)})$ be a sequence of parameters satisfying (3.3). We shall show that the distribution of $\sqrt{J}(\widehat{\theta}_J^{(m)}-\theta_{(J)})$, under $(\theta_{(J)},h_{(J)})$, converges weakly to $N(0,(V_m(t_0))^{-1})$ which depends on $(\theta_{(0)},h_{(0)})$, but not on the particular sequence $\{(\theta_{(J)},h_{(J)}\}$.

Let

$$\widetilde{G}_J(\theta_{(0)}, h_{(0)}, \delta, \beta, t) \equiv \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \sum_{i=0}^{\infty}\left(\frac{S_{iJ}^{(1)}(\theta_{(0)}, u)}{S_{iJ}^{(0)}(\theta_{(0)}, u)} + \frac{\beta_i(u)}{\delta h_{(0),i}(u)}\right) 1_{(T_{ji}, T_{ji+1}]}(u)\, dM_{jk}(u).$$

We note that $\widetilde{G}_J(\theta_{(0)}, h_{(0)}, \delta, \beta, \cdot)$ is a square-integrable martingale and is orthogonal to $G_J^{(2)}(\theta_{(0)}, \cdot)$. This together with some relevant boundedness condition shows that, by martingale central limit theorem [cf. Andersen and Gill (1982)], the random vector

$$\left(\frac{1}{\sqrt{J}} G_J^{(2)}(\theta_{(0)}, t_0), \frac{1}{\sqrt{J}}\widetilde{G}_J(\theta_{(0)}, h_0, \delta, \beta, t_0)\right)$$

converges weakly to a Gaussian vector $(Z, \widetilde{Z})$, where Z and $\widetilde{Z}$ are two independent mean-zero random variables with variances $V_2(t_0)$ and $\widetilde{V}(t_0)$, respectively, with

$$\widetilde{V}(t) = \sum_{i=0}^{\infty}\left(\frac{s_i^{(1)}(\theta_{(0)}, u)}{s_i^{(0)}(\theta_{(0)}, u)} + \frac{\beta_i(u)}{\delta h_{(0),i}(u)}\right)^2 h_{(0),i}(u) s_i^{(0)}(\theta_{(0)}, u)\, du. \tag{5.2}$$

Let

$$\widetilde{\mathcal{L}}_J(t_0) = \mathcal{L}_J(t_0, \theta_{(J)}, h_{(J)}) - \mathcal{L}_J(t_0, \theta_{(0)}, h_{(0)}).$$

It follows from straightforward calculation that

$$\begin{aligned}
\widetilde{\mathcal{L}}_J(t) = {} & \frac{\delta}{\sqrt{J}}\sum_{j=1}^{K}\sum_{k=1}^{K}\int_0^t \left(Z_{jk}(u) - \sum_{i=0}^{\infty}\frac{S_{iJ}^{(1)}(\theta_{(0)}, u)}{S_{iJ}^{(0)}(\theta_{(0)}, u)} 1_{(T_{ji}, T_{ji+1}]}(u)\right) dM_{jk}(u) \\
& + \frac{\delta}{\sqrt{J}}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^t \sum_{i=0}^{\infty}\left(\frac{S_{iJ}^{(1)}(\theta_{(0)}, u)}{S_{iJ}^{(0)}(\theta_{(0)}, u)} + \frac{\beta_i(u)}{\delta h_{(0),i}(u)}\right) 1_{(T_{ji}, T_{ji+1}]}(u)\, dM_{jk}(u) \\
& - \frac{\delta^2}{2}\sum_{i=0}^{\infty}\int_0^t h_{(0),i}(u) S_{iJ}^{(2)}(\theta_{(0)}, u)\, du - \frac{1}{2}\sum_{i=0}^{\infty}\int_0^t \frac{(\beta_i(u))^2}{h_{(0),i}(u)} S_{iJ}^{(0)}(\theta_{(0)}, u)\, du \qquad (5.3)\\
& - \delta\sum_{i=0}^{\infty}\int_0^t \beta_i(u) S_{iJ}^{(1)}(\theta_{(0)}, u)\, du + R_J(t) \\
= {} & \frac{\delta}{\sqrt{J}} G_J^{(2)}(\theta_{(0)}, t) + \frac{\delta}{\sqrt{J}}\widetilde{G}_J(\theta_{(0)}, h_{(0)}, \delta, \beta, t) \\
& - \frac{\delta^2}{2}\sum_{i=0}^{\infty}\int_0^t h_{(0),i}(u)\left(S_{iJ}^{(2)}(\theta_{(0)}, u) - \frac{\left(S_{iJ}^{(1)}(\theta_{(0)}, u)\right)^2}{S_{iJ}^{(0)}(\theta_{(0)}, u)}\right) du \qquad (5.4)\\
& - \frac{\delta^2}{2}\sum_{i=0}^{\infty}\int_0^t \left(\frac{S_{iJ}^{(1)}(\theta_{(0)}, u)}{S_{iJ}^{(0)}(\theta_{(0)}, u)} + \frac{\beta_i(u)}{\delta h_{(0),i}(u)}\right)^2 h_{(0),i}(u) S_{iJ}^{(0)}(\theta_{(0)}, u)\, du + R_J(t),
\end{aligned}$$

where $\sup_{0 \le t \le t_0} |R_J(t)|$ converges to 0 in probability.

We note also that

$$\sqrt{J}\left(\widehat{\theta}_J^{(2)} - \theta_{(0)}\right) = -\frac{1}{\sqrt{J}} G_J^{(2)}(\theta_{(0)}, t_0) \left(\frac{1}{J} \frac{\partial}{\partial \theta} G_J^{(2)}(\theta_J^*, t_0) \right)^{-1}, \tag{5.5}$$

where θ_J^* is on the line segment between $\widehat{\theta}_J^{(2)}$ and $\theta_{(0)}$.

The characteristic function of $\sqrt{J}(\widehat{\theta}_J^{(2)} - \theta_{(J)})$, under $(\theta_{(J)}, h_{(J)})$, can be written as

$$\begin{aligned} &E_{(\theta_{(J)}, h_{(J)})} \exp\left[iu\sqrt{J}\left(\widehat{\theta}_J^{(2)} - \theta_{(J)}\right)\right] \\ &\quad = E_{(\theta_{(J)}, h_{(J)})} \exp\left[iu\sqrt{J}\left(\widehat{\theta}_J^{(2)} - \theta_{(0)}\right) - iu\sqrt{J}\left(\theta_{(J)} - \theta_{(0)}\right)\right] \\ &\quad = E_{(\theta_{(0)}, h_{(0)})} \exp\left[iu\sqrt{J}\left(\widehat{\theta}_J^{(2)} - \theta_{(0)}\right) - iu\sqrt{J}\left(\theta_{(J)} - \theta_{(0)}\right) + \widetilde{\mathcal{L}}_J(t_0)\right]. \end{aligned} \tag{5.6}$$

It follows from (5.2), (5.4) and weak convergence of $G_J^{(2)}$ and $\widetilde{G}_J$ that $\widetilde{\mathcal{L}}_J(t_0)$ converges weakly to $\delta Z + \delta\widetilde{Z} - (\delta^2/2)V_2(t_0) - (\delta^2/2)\widetilde{V}(t_0)$, where Z and $\widetilde{Z}$ are given above. Since $E_{(\theta_{(0)}, h_{(0)})} \exp[\widetilde{\mathcal{L}}_J(t_0)] = 1$, $\{\exp[\widetilde{\mathcal{L}}_J(t_0)] \mid J = 1, 2, \ldots\}$ is uniformly integrable [cf. Billingsley (1968), Theorem 5.4]. This together with (5.5) implies that (5.6) converges to

$$\begin{aligned} &E \exp\left[iu\left(V_2(t_0)\right)^{-1} Z - iu\delta + \delta Z - \frac{\delta^2}{2} V_2(t_0) + \delta\widetilde{Z} - \frac{\delta^2}{2}\widetilde{V}(t_0)\right] \\ &\quad = E \exp\left[\left(iu\left(V_2(t_0)\right)^{-1} + \delta\right) Z - iu\delta - \frac{\delta^2}{2} V_2(t_0)\right] E \exp\left[\delta\widetilde{Z} - \frac{\delta^2}{2}\widetilde{V}(t_0)\right] \\ &\quad = E \exp\left[-\frac{u_2}{2V_2(t_0)} + iu\delta + \frac{\delta^2}{2} V_2(t_0) - iu\delta - \frac{\delta^2}{2} V_2(t_0)\right] \\ &\qquad \times \exp\left[\frac{\delta^2}{2}\widetilde{V}(t_0) - \frac{\delta^2}{2}\widetilde{V}(t_0)\right] \\ &\quad = \exp\left[-\frac{u_2}{2V_2(t_{(0)})}\right]. \end{aligned} \tag{5.7}$$

This completes the proof. □

6. Concluding remarks.

1. This paper studies the estimation problem with the self-excited filtration of the counting processes. In fact, according to a recent work of Greenwood and Welfelmeyer (1993), the main results of this paper seem to hold true also when more general filtrations are used so that the covariates can be extraneous to the counting processes. The main idea for this extension is

that we require (3.2) to be the only statistically "relevant" part of the log-likelihood process with respect to the filtration considered.

2. In order to avoid technical complications, we imposed many stringent regularity conditions. For example, the boundedness assumption about Y_j, Z_j, t_0, Θ and the nuisance parameter simplifies the argument in many places. We worked basically with L_2-theory. Many of the results here can be generalized directly to include a larger class of baseline hazard rates so that the martingales involved are locally square-integrable martingales. Obviously, the results of this paper can also be extended to the case for which Θ is in R_d, $d > 1$.
3. In this paper, we study asymptotics as the number of replicates of the model, J, gets large. There are other possibilities. We may, following Greenwood and Wefelmeyer (1991), study the asymptotics as t gets large for fixed J and K; or we may consider the case when K is large. These and other related issues will be treated in a future study.
4. We note that, by satisfying the data according to the number of previous occurrences, we obtain many "smaller" proportional hazards models in both Model 2 and Model 3. Estimating functions considered for these two models are sums of estimating functions for the "smaller" proportional hazards models. In Model 2, this is done using natural calendar time filtration; in Model 3, we find it successful when a different filtration is used. Identifying this useful filtration is an important part of the work. It would be interesting if these two models can be given a unified formal treatment.

Acknowledgments. The authors are grateful for the suggestions and comments made by the referees and an Associate Editor.

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.

BEGUN, J. M., HALL, W. J., HUANG, W. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11** 432–453.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

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

CHANG, I. S. and HSIUNG, C. A. (1991a). Applications of estimating function theory to replicates of generalized proportional hazards models. In *Estimating Functions* (V. P. Godambe, ed.) 23–33. Oxford Univ. Press.

CHANG, I. S. and HSIUNG, C. A. (1991b). An E-ancillarity projection property of Cox's partial score function. *Ann. Statist.* **19** 1651–1660.

CHANG I. S. and HSIUNG, C. A. (1991c). A large sample study of some generalized proportional hazards models. Technical report, Inst. Statistical Science, Academia Sinica, Taipei.

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

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

ELLIOTT, R. J. (1982). *Stochastic Calculus and Applications.* Springer, New York.

GAIL, M. H., SANTNER, T. J. and BROWN, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36** 255–266.

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

GREENWOOD, P. E. and WEFELMEYER, W. (1991). On optimal estimating functions for partially specified counting process models. In *Estimating Functions* (V. P. Godambe ed.) 147–160. Oxford Univ. Press.

GREENWOOD, P. E. and WEFELMEYER, W. (1993). Private communications.

MILLAR, P. W. (1989). Optimal estimation in the non-parametric multiplicative intensity model. Technical Report 173, Dept. Statistics, Univ. California, Berkeley.

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

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales, and the Cox model. In *Statistical Inference from Stochastic Processes* (N. U. Prabhu, ed.) 191–220. Amer. Math. Soc., Providence, RI.

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

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