The Annals of Statistics
1995, Vol. 23, No. 5, 1712–1734

SEMIPARAMETRIC ANALYSIS OF GENERAL ADDITIVE–MULTIPLICATIVE HAZARD MODELS FOR COUNTING PROCESSES

BY D. Y. LIN[1] AND ZHILIANG YING[2]

University of Washington and Rutgers University

The additive–multiplicative hazard model specifies that the hazard function for the counting process associated with a multidimensional covariate process $Z = (W^T, X^T)^T$ takes the form of $\lambda(t \mid Z) = g\{\beta_0^T W(t)\} + \lambda_0(t)h\{\gamma_0^T X(t)\}$, where $\theta_0 = (\beta_0^T, \gamma_0^T)^T$ is a vector of unknown regression parameters, g and h are known link functions and λ_0 is an unspecified "baseline hazard function." In this paper, we develop a class of simple estimating functions for θ_0, which contains the partial likelihood score function in the special case of proportional hazards models. The resulting estimators are shown to be consistent and asymptotically normal under appropriate regularity conditions. Weak convergence of the Aalen–Breslow type estimators for the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\,du$ is also established. Furthermore, we construct adaptive estimators for θ_0 and Λ_0 that achieve the (semiparametric) information bounds. Finally, a real example is provided along with some simulation results.

1. Introduction. Semiparametric regression models based on the hazard function or intensity process provide a natural and convenient framework for studying the influence of the covariate history on the (possibly censored) failure time or the general counting process. The most familiar hazard-based formulation is the Cox (1972) proportional hazards model or the multiplicative hazard model, which assumes that the hazard function associated with a multidimensional covariate process $X(\cdot)$ is

$$\lambda(t \mid X) = \lambda_0(t)\exp\left[\gamma_0^T X(t)\right], \tag{1.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function and γ_0 is an unknown parameter vector. A plausible alternative is the additive hazard model in the form of

$$\lambda(t \mid X) = \lambda_0(t) + \beta_0^T X(t), \tag{1.2}$$

where β_0 is an unknown parameter vector [Cox and Oakes (1984), page 74, Thomas (1986), Breslow and Day (1987), page 182, and Lin and Ying (1994)].

Received November 1992; revised August 1993.
[1] Supported by NIH Grant R29 GM-47845.
[2] Supported by NSF Grant DMS-92-07730 and NSA Grant MDA-904-92-H-3058.
AMS 1991 *subject classifications*. Primary 62J99; secondary 62P10.
Key words and phrases. Aalen–Breslow estimator, adaptive estimation, asymptotic efficiency, censoring, Cox regression, estimating equation, failure time, information bound, martingale, partial likelihood, proportional hazards, survival data, time-dependent covariate.

The exponential link function in (1.1) and the linear link function in (1.2) may be replaced by other forms.

Analysis of model (1.1) is normally based on the partial likelihood principle [Cox (1972, 1975)]. The maximum partial likelihood estimator has been shown to be consistent, asymptotically normal and asymptotically efficient [Tsiatis (1981), Andersen and Gill (1982) and Andersen, Borgan, Gill and Keiding (1993), pages 481–509, 647–650]. In addition, Prentice and Self (1983) provided a large-sample theory for the partial likelihood analysis of the multiplicative hazard model with a general link function.

The partial likelihood function for model (1.2), however, is intractable. Recently, Lin and Ying (1994) presented an explicit estimator for β_0 resulting from an analog of the partial likelihood score equation for γ_0. In related developments, Aalen (1980, 1989), Huffer and McKeague (1991) and Andersen, Borgan, Gill and Keiding [(1993), VII.4] described least-squares type estimators for Aalen's nonparametric additive hazard model $\lambda(t \mid X) = \zeta(t)^T X(t)$ [Aalen (1980)] which is less parsimonious than model (1.2). Furthermore, McKeague and Sasieni (1994) studied an additive model that is a mixture of model (1.2) and Aalen's model.

The additive and multiplicative hazard models postulate two rather different relationships between the covariate process and the hazard function. In some applications physical rationale dictates that one model is more adequate than the other, whereas in others the choice between the two models is an empirical matter. The two models can hold simultaneously if $\lambda_0(\cdot)$ is time-invariant, in which case it is desirable to estimate the regression parameters under both models as the relative risk and risk difference are complementary measures. To enhance our modeling capability, it seems natural to consider models which allow some covariate effects to be additive while allowing others to be multiplicative or which allow certain covariates to have both the additive and multiplicative effects. We therefore study the following class of general additive–multiplicative hazard models:

$$\lambda(t \mid Z) = g\{\beta_0^T W(t)\} + \lambda_0(t) h\{\gamma_0^T X(t)\}, \tag{1.3}$$

where $Z = (W^T, X^T)^T$ is a p-vector of covariates, $\theta_0 = (\beta_0^T, \gamma_0^T)^T$ is a p-vector of unknown regression parameters, g and h are known link functions and λ_0 is an unspecified "baseline hazard function" under $g = 0$ and $h = 1$. Obviously, (1.3) encompasses both models (1.1) and (1.2). Some common examples of the link function h are $h(x) = e^x$ and $h(x) = 1 + x$ and those of g are $g(x) = x$ and $g(x) = e^x$. For $g(x) = e^x$, it is sensible to let the first component of W be 1.

In the next section of this paper, we develop a class of estimating functions for θ_0, which resembles the partial likelihood score function under model (1.1). The resulting estimators are shown to be consistent and asymptotically normal under broad conditions. In Section 3, we establish the asymptotic properties of the Aalen–Breslow type estimators for the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$. Section 4 presents our simulation results and a real-life example. In Section 5, we construct adaptive estimators for θ_0 and Λ_0 that are asymptotically efficient.

2. Estimation of θ_0. In this section, we shall formulate model (1.3) in the framework of multivariate counting processes and develop some simple estimators for θ_0. We shall be working on the finite time interval $[0, \tau]$, although the extension to infinite τ will be discussed in Section 2.4. We shall use basic results from the theory of multivariate counting processes and martingales as surveyed in Andersen, Borgan, Gill and Keiding (1993) without further comment.

2.1. *Formulation of the model.* Let $N_1(t), \dots, N_n(t)$ be n independent counting processes adapted to a filtration $\{\mathscr{F}_t: t \in [0, \tau]\}$ satisfying the usual conditions as described in Andersen, Borgan, Gill and Keiding [(1993), II.4]. We allow each N_i to take multiple jumps, although in the classical survival setting there is at most one jump for each subject. Associated with each $N_i(t)$ is a (p-dimensional) covariate process $Z_i(t) = \{W_i^T(t), X_i^T(t)\}^T$ that is $\mathscr{F}_t$-predictable. Suppose that the compensator for $N_i(t)$ takes the form $\int_0^t Y_i(s)\lambda(s \mid Z_i)\,ds$, where $\lambda(t \mid Z)$ is given by (1.3) and $Y_i(t)$ is a $\{0, 1\}$-valued left-continuous process adapted to $\mathscr{F}_t$. Thus,

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\left[g\{\beta_0^T W_i(s)\} + h\{\gamma_0^T X_i(s)\}\lambda_0(s)\right] ds$$

is an $\mathscr{F}_t$-martingale. Furthermore, we assume throughout Section 2 that there exists $\delta_\tau > 0$ such that

$$\left[n^{-1} \inf_{t \le \tau;\, \|\gamma - \gamma_0\| \le \delta_\tau} \sum_{i=1}^n Y_i(t) h\{\gamma^T X_i(t)\}\right]^{-1} = O_p(1).$$

2.2. *Estimating functions for θ_0.* If the baseline hazard function λ_0 is known, then the likelihood for $\theta_0 = (\beta_0^T, \gamma_0^T)^T$ is proportional to

$$\prod_{i=1}^n \left[\left\{\prod_{t \le \tau} \lambda(t \mid Z_i)^{dN_i(t)}\right\} \exp\left\{-\int_0^\tau Y_i(t)\,\lambda(t \mid Z_i)\,dt\right\}\right]$$

[Andersen, Borgan, Gill and Keiding (1993), pages 58–59]. The corresponding (p-dimensional) score function is

$$\sum_{i=1}^n \int_0^\tau \frac{g'\{\beta^T W_i(t)\} W_i(t)}{g\{\beta^T W_i(t)\} + h\{\gamma^T X_i(t)\}\lambda_0(t)}\, dM_i(\theta, t), \tag{2.1}$$

$$\sum_{i=1}^n \int_0^\tau \frac{h'\{\gamma^T X_i(t)\} X_i(t)\, \lambda_0(t)}{g\{\beta^T W_i(t)\} + h\{\gamma^T X_i(t)\}\lambda_0(t)}\, dM_i(\theta, t), \tag{2.2}$$

where

$$M_i(\theta,t) = N_i(t) - \int_0^t Y_i(s)\left[g\{\beta^T W_i(s)\} + h\{\gamma^T X_i(s)\}\lambda_0(s)\right] ds.$$

Since $M_i(\theta_0,t) = M_i(t)$, the above score function is a martingale integral at $\theta = \theta_0$.

In order to develop inference procedures for θ_0 with λ_0 being completely unspecified, we shall modify (2.1) and (2.2) to eliminate λ_0 from the estimating functions. First, we replace the integrands in (2.1) and (2.2) by a (p-dimensional) predictable process, denoted by $D_i(\theta,t)$, which is a smooth function of Z_i and θ not involving λ_0. Second, we replace $\lambda_0(t)\,dt$ in $M_i(\theta, t)$ by $d\hat{\Lambda}_0(\theta, t)$, where $\hat{\Lambda}_0(\theta_0, t)$ is the Aalen–Breslow type estimator for $\Lambda_0(t)$ with known θ_0, namely,

$$\hat{\Lambda}_0(\theta_0,t) = \int_0^t \frac{\sum_{i=1}^n\left[dN_i(s) - Y_i(s)g\{\beta_0^T W_i(s)\}\,ds\right]}{\sum_{i=1}^n Y_i(s)h\{\gamma_0^T X_i(s)\}}. \tag{2.3}$$

We then obtain an ad hoc (p-dimensional) estimating function $S(\theta, \tau)$ free of λ_0, where

$$\begin{aligned} S(\theta,t) &= \sum_{i=1}^n \int_0^t D_i(\theta,s)\big[dN_i(s) - Y_i(s)g\{\beta^T W_i(s)\}\,ds \\ &\qquad\qquad -Y_i(s)h\{\gamma^T X_i(s)\}\,d\hat{\Lambda}_0(\theta,s)\big] \\ &= \sum_{i=1}^n \int_0^t \{D_i(\theta,s) - \overline{D}(\theta,s)\}\left[dN_i(s) - Y_i(s)g\{\beta^T W_i(s)\}\,ds\right], \end{aligned}$$

with $\overline{D}(\theta, t) = \sum_{i=1}^n Y_i(t)h\{\gamma^T X_i(t)\}D_i(\theta, t)/\sum_{i=1}^n Y_i(t)h\{\gamma^T X_i(t)\}$.

Since $\sum_{i=1}^n \{D_i(\theta, s) - \overline{D}(\theta, s)\}Y_i(s)h\{\gamma^T X_i(s)\} = 0$, we observe that

$$S(\theta,t) = \sum_{i=1}^n \int_0^t \{D_i(\theta,s) - \overline{D}(\theta,s)\}\,dM_i(\theta,s),$$

which indicates that $S(\theta_0,t)$ is a martingale. The fact that $S(\theta_0,t)$ is a martingale is not a great stroke of algebraic luck, but is instead a natural consequence of the underlying mathematical structure in our construction. First, the replacement of the integrand in a martingale integral by a different predictable process does not destroy the martingale nature. Second,

$$d\hat{\Lambda}_0(\theta_0,t) - d\Lambda_0(t) = \frac{\sum_{i=1}^n dM_i(t)}{\sum_{i=1}^n Y_i(t)h\{\gamma_0^T X_i(t)\}}$$

provided that $\sum_{i=1}^n Y_i(t)h\{\gamma_0^T X_i(t)\} > 0$; in fact, (2.3) was derived from the martingale equation $\sum dM_i(t) = \sum dN_i(t) - \sum Y_i(t)[g\{\beta_0^T W_i(t)\}\,dt + h\{\gamma_0^T X_i(t)\}\,d\Lambda_0(t)]$ by setting the noise to 0. Therefore, the replacement of $d\Lambda_0(t)$ in $M_i(\theta, t)$ by $d\hat{\Lambda}_0(\theta, t)$ will yield a different martingale integral at $\theta = \theta_0$. The martingale representations for $S(\theta_0, t)$ and $\hat{\Lambda}_0(\theta_0, t)$ will enable us to establish weak convergence and simplify variance calculation.

The following assumptions are required for subsequent developments.

STABILITY CONDITION 1. There exists an integrable function v such that, for every $t \in [0, \tau]$,

$$n^{-1} \sum_{i=1}^{n} Y_i(t)\left[g\{\beta_0^T W_i(t)\} + h\{\gamma_0^T X_i(t)\}\lambda_0(t)\right] \times \{D_i(\theta_0, t) - \bar{D}(\theta_0, t)\}^{\otimes 2} \to_P v(t).$$

Here and in the sequel, $a^{\otimes 2}$ denotes the out-product aa^T of a column vector a.

NEGLIGIBILITY CONDITION. For any $\varepsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \|D_i(\theta_0, s) - \bar{D}(\theta_0, s)\|^2 1_{\{\|D_i(\theta_0,s)-\bar{D}(\theta_0,s)\|^2 > n\varepsilon\}} Y_i(s) \times \left[g\{\beta_0^T W_i(s)\} + h\{\gamma_0^T X_i(s)\}\lambda_0(s)\right] ds \to_P 0.$$

REMARK 2.1. The above two conditions are analogous to the variance–covariance stability and Lindeberg conditions of the classical multivariate central limit theorem for sum of independent zero-mean random vectors. The negligibility condition is clearly satisfied when $\{D_i(\theta_0, t)\}$ are bounded uniformly in i and t. Further discussions will be provided in Section 2.4.

THEOREM 2.1. *Suppose that stability condition* 1 *and the negligibility condition hold. Then* $n^{-1/2}S(\theta_0, \cdot)$ *converges weakly in* $\mathscr{D}[0, \tau]$ *to a zero-mean Gaussian process with independent increments and with variance function* $V_t = \int_0^t v(s)\, ds$.

PROOF. As stated above, $n^{-1/2}S(\theta_0, \cdot)$ is a martingale integral. It suffices to verify conditions (2.5.1) and (2.5.3) of Andersen, Borgan, Gill and Keiding [(1993), Theorem II.5.1 (Rebolledo's central limit theorem)]. The second condition is clearly implied by our negligibility condition. The first one is also satisfied since the predictable variation process of $n^{-1/2}S(\theta_0, t)$ is

$$n^{-1} \sum_{i=1}^{n} \int_0^{t} \{D_i(\theta_0, s) - \bar{D}(\theta_0, s)\}^{\otimes 2} \times Y_i(s)\left[g\{\beta_0^T W_i(s)\} + h\{\gamma_0^T X_i(s)\}\lambda_0(s)\right] ds,$$

which converges in probability to V_t by stability condition 1. □

2.3. *Asymptotic properties of* $\hat{\theta}$. In light of Theorem 2.1, we can use S to estimate θ_0. Specifically, define $\hat{\theta}$ as a root to $S(\theta, \tau) = 0$. We make the following assumption:

STABILITY CONDITION 2. There exists a $p \times p$ matrix A such that

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t)\{D_i(\theta_0, t) - \bar{D}(\theta_0, t)\}\left[g'\{\beta_0^T W_i(t)\} W_i^T(t),\ h'\{\gamma_0^T X_i(t)\} X_i^T(t)\lambda_0(t)\right] dt \to_P A.$$

THEOREM 2.2. *Suppose that the negligibility condition and stability conditions* 1 *and* 2 *are satisfied. Suppose also that A is nonsingular and that the family of partial derivatives (with respect to θ) of $D_i(\theta, t)$, $g\{\beta^T W_i(t)\}$ and $h\{\gamma^T X_i(t)\}$, as functions of θ, is equicontinuous in a neighborhood of θ_0. Then there exists a neighborhood of θ_0 within which, with probability tending to* 1 *as $n \to \infty$, the root $\hat\theta$ of $S(\theta, \tau) = 0$ is uniquely defined and $n^{1/2}(\hat\theta - \theta_0) \to_L \mathcal{N}(0, A^{-1}V_\tau A^{-1^T})$.*

PROOF. The Taylor expansion of $S(\hat\theta, \tau)$ around θ_0 yields

$$n^{1/2}\left(\hat\theta - \theta_0\right) = \left\{-n^{-1}\frac{\partial S(\theta^*, \tau)}{\partial\theta}\right\}^{-1} n^{-1/2}S(\theta_0, \tau),$$

where θ^* is on the line segment between $\hat\theta$ and θ_0. Thus, by Theorem 2.1, $n^{1/2}(\hat\theta - \theta_0) \to_L \mathcal{N}(0, A^{-1}V_\tau A^{-1^T})$ provided that $\hat\theta \to_P \theta_0$ and $-n^{-1} \times \partial S(\theta_0, \tau)/\partial\theta \to_P A$, which are verified below.

We first show that $-n^{-1}\ \partial S(\theta_0, \tau)/\partial\theta \to_P A$. Clearly,

$$\begin{aligned} -n^{-1}\frac{\partial S(\theta_0, \tau)}{\partial\theta} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\{D_i(\theta_0, t) - \bar D(\theta_0, t)\} \\ &\quad \times \frac{\partial\left[g\{\beta_0^T W_i(t)\} + h\{\gamma_0^T X_i(t)\}\lambda_0(t)\right]}{\partial\theta}\,dt \\ &\quad - n^{-1}\sum_{i=1}^{n}\int_0^\tau\left[\frac{\partial\{D_i(\theta_0, t) - \bar D(\theta_0, t)\}}{\partial\theta}\right] dM_i(t). \end{aligned} \tag{2.4}$$

Stability condition 2 implies that the first term on the right-hand side of (2.4) converges to A. The second term is $o_p(1)$ because it is a martingale integral whose predictable variation at τ is $O_p(n^{-1}) = o_p(1)$, noting that the summation is scaled by n^{-1} rather than the usual $n^{-1/2}$.

From the equicontinuity assumptions, it follows that for any $\varepsilon > 0$ we can choose $\delta > 0$ such that, for all n, $\|n^{-1}\ \partial S(\theta, \tau)/\partial\theta - n^{-1}\ \partial S(\theta_0, \tau)/\partial\theta\| < \varepsilon$ whenever $\|\theta - \theta_0\| \le \delta$. This result together with the fact that $-n^{-1} \times \partial S(\theta_0, \tau)/\partial\theta \to_P A$ implies that

$$P\left\{\sup_{\|\theta - \theta_0\| \le \delta}\left\|-n^{-1}\frac{\partial S(\theta, \tau)}{\partial\theta} - A\right\| > 2\varepsilon\right\} \to 0 \quad \text{as } n \to \infty. \tag{2.5}$$

Now, Theorems 4.1 and 4.2 of Goffman [(1965), pages 89–91] assert that if $f(x)$, $x \in R^p$, is continuously differentiable at x_0 and $\partial f(x_0)/\partial x$ is nonsingular, then there exist δ_0 and ε_0 such that f is a one-to-one mapping on $B(x_0; \delta_0)$, the ball centered at x_0 with radius δ_0, and $f(B(x_0; \delta_0)) \supset B(f(x_0); \varepsilon_0)$. Furthermore, it is evident from the proofs of the two theorems that the results hold simultaneously for a family of such functions with common δ_0 and ε_0 as long as their derivatives at x_0 are sufficiently close. In view of these results and (2.5), we conclude that there exist δ_1 and ε_1 such that $n^{-1}S$ is a one-to-one mapping from the $B(\theta_0; \delta_1)$ to

$n^{-1}S(B(\theta_0;\, \delta_1), \tau)$, which contains $B(n^{-1}S(\theta_0, \tau);\, \varepsilon_1)$. Since $n^{-1} \times S(\theta_0, \tau) \to_P 0$, $B(n^{-1}S(\theta_0, \tau);\, \varepsilon_1)$ contains 0 for all large n; therefore, $\hat{\theta}$ exists and is unique in $B(\theta_0;\, \delta_1)$, and $\hat{\theta} \to_P \theta_0$. □

Theorem 2.2 only ensures the existence and uniqueness of the root of S locally. The following result pertains to the global uniqueness.

COROLLARY 2.3. *In addition to the assumptions of Theorem 2.2, suppose that the equicontinuity condition in Theorem 2.2 is satisfied on a compact region G whose interior contains θ_0 and $n^{-1}S(\theta, \tau) \to_P \psi(\theta)$ for each $\theta \in G$. Then, with probability tending to 1 as $n \to \infty$, the estimating function $S(\theta, \tau)$ has a unique root in G if θ_0 is the only root of $\psi(\theta)$.*

PROOF. Because of the boundedness (uniformly in θ) of the derivative of $n^{-1}S(\theta, \tau)$, the convergence of $n^{-1}S(\theta, \tau)$ to $\psi(\theta)$ is uniform on G. Thus, with probability tending to 1, $n^{-1}S(\theta, \tau)$ stays away from 0 outside any fixed neighborhood of θ_0. Combining this result with Theorem 2.2, we get the desired global asymptotic uniqueness. □

The assumption that θ_0 is the only root of $\psi(\theta)$ is satisfied in the special cases of models (1.1) and (1.2). The following example illustrates that the global uniqueness also holds in other nontrivial settings.

EXAMPLE 2.1. Suppose that the counting processes are generated by the failure times T_i, that is, $N_i(t) = 1_{\{T_i \le t\}}$ and $Y_i(t) = 1_{\{T_i \ge t\}}$. Suppose also that (T_i, W_i, X_i) are jointly i.i.d. and W_i and X_i are independent. In addition, let $g(x) = x$, $h(x) = e^x$ and $D_i^T = (W_i^T, X_i^T)$. For notational simplicity, assume that W_i and X_i are one-dimensional and time-independent. Finally, let $\overline{D}_k$, S_k and ψ_k, $k = 1, 2$, denote the kth components of $\overline{D}$, S and ψ. In our setting, $E\{Y_i(t) \mid Z_i\} = \exp(-\beta_0 W_i t)\exp\{-\exp(\gamma_0 X_i)\Lambda_0(t)\}$. This entails that $\overline{D}_1(\theta, t)$ converges to $\bar{d}_1(t) = E\{W_1 \exp(-\beta_0 W_1 t)\}/E\{\exp(-\beta_0 W_1 t)\}$, which involves neither β nor γ. Thus, $\psi_1(\theta)$ is a linear function of β only and therefore has a unique root, which must be β_0. Hence, to show that $\psi(\theta)$ has a unique root, it suffices to show that $\psi_2(\beta_0, \gamma)$ has a unique root. Note that

$$S_2(\beta_0, \gamma; \tau) = \sum_{i=1}^{n} \int_0^{\tau} \{X_i - \overline{D}_2(\gamma, t)\}\{dN_i(t) - Y_i(t)\beta_0 W_i\, dt\},$$

where $\overline{D}_2$ does not involve β. The fact that $dM_i(t) = dN_i(t) - Y_i(t)\beta_0 W_i\, dt - Y_i(t)\exp(\gamma_0 X_i)\lambda_0(t)\, dt$ implies that $n^{-1}S_2(\beta_0, \gamma; \tau)$ has the same limit as

$$n^{-1}\tilde{S}_2(\beta_0, \gamma) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \{X_i - \overline{D}_2(\gamma, t)\} Y_i(t) \exp(\gamma_0 X_i)\lambda_0(t)\, dt.$$

It is easy to show that

$$-\frac{\partial \tilde{S}_2(\beta_0,\gamma)}{\partial \gamma} = \int_0^{\tau} \frac{\sum_{i=1}^n \{X_i - \overline{D}_2(\gamma,t)\}^2 \exp(\gamma X_i) Y_i(t)}{\sum_{j=1}^n \exp(\gamma X_j) Y_j(t)} \times \sum_{l=1}^n Y_l(t) \exp(\gamma_0 X_l) \lambda_0(t)\, dt,$$

which is always nonnegative. Since $\psi_2(\beta_0,\gamma_0) = 0$ and $\partial\psi_2(\beta_0, \gamma_0)/\partial\gamma \neq 0$, γ_0 must be the unique root.

REMARK 2.2. In practice, whenever there are multiple roots for $S(\theta, \tau)$, we face the problem of choosing the right one. A convenient solution is to construct another estimating function, say, $S^*(\theta, \tau)$, using a different set of $\{D_i\}$ and then to choose the root of $S(\theta, \tau)$ which makes $S^*(\theta, \tau)$ closest to 0. The rationale is that both $S(\theta, \tau)$ and $S^*(\theta, \tau)$ have mean 0 at θ_0, so the root identified by the above criterion is likely to be the consistent one.

It is natural to estimate the limiting covariance matrix of $n^{1/2}(\hat{\theta} - \theta_0)$ by $\hat{A}^{-1}(\hat{\theta})\hat{V}(\hat{\theta})\hat{A}^{-1}(\hat{\theta})^T$, where

$$\hat{A}(\theta) = n^{-1} \sum_{i=1}^n \int_0^{\tau} Y_i(t)\{D_i(\theta,t) - \overline{D}(\theta,t)\} \times \left[g'\{\beta^T W_i(t)\} W_i^T(t)\, dt, h'\{\gamma^T X_i(t)\} X_i^T(t)\, d\hat{\Lambda}_0(\theta,t)\right],$$

$$\hat{V}(\theta) = n^{-1} \sum_{i=1}^n \int_0^{\tau} \{D_i(\theta,t) - \overline{D}(\theta,t)\}^{\otimes 2}\, dN_i(t).$$

Because the covariance matrix estimator involves $\hat{\Lambda}_0(\hat{\theta}, \cdot)$, we shall defer the proof of its consistency to Section 3 after establishing the asymptotic properties of $\hat{\Lambda}_0$.

2.4. *Some special cases.* One possible choice of $D_i(\theta, t)$ is

$$\begin{bmatrix} \tilde{W}_i(\theta,t) \\ \tilde{X}_i(\theta,t) \end{bmatrix} = \begin{bmatrix} g'\{\beta^T W_i(t)\} W_i(t)/h\{\gamma^T X_i(t)\} \\ h'\{\gamma^T X_i(t)\} X_i(t)/h\{\gamma^T X_i(t)\} \end{bmatrix}.$$

Obviously, $\tilde{W}_i$ and $\tilde{X}_i$ will be good approximations to the integrands in (2.1) and (2.2) if λ_0 is roughly constant and g is small relative to $h\lambda_0$. Here,

$$S(\theta,\tau) = \begin{bmatrix} U_\beta(\theta) \\ U_\gamma(\theta) \end{bmatrix},$$

where

$$U_\beta(\theta) = \sum_{i=1}^n \int_0^{\tau} \{\tilde{W}_i(\theta,t) - \overline{W}(\theta,t)\}\left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right],$$

$$U_\gamma(\theta) = \sum_{i=1}^n \int_0^{\tau} \{\tilde{X}_i(\theta,t) - \overline{X}(\theta,t)\}\left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right],$$

with

$$\overline{W}(\theta,t) = \frac{\Sigma Y_i(t)h\{\gamma^T X_i(t)\}\tilde{W}_i(\theta,t)}{\Sigma Y_i(t)h\{\gamma^T X_i(t)\}}$$

and

$$\overline{X}(\theta,t) = \frac{\Sigma Y_i(t)h\{\gamma^T X_i(t)\}\tilde{X}_i(\theta,t)}{\Sigma Y_i(t)h\{\gamma^T X_i(t)\}}.$$

In addition,

$$\hat{A}(\theta) = \begin{bmatrix} \hat{A}_{\beta\beta}(\theta) & \hat{A}_{\beta\gamma}(\theta) \\ \hat{A}_{\gamma\beta}(\theta) & \hat{A}_{\gamma\gamma}(\theta) \end{bmatrix},$$

where

$$\hat{A}_{\beta\beta}(\theta) = n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)h\{\gamma^T X_i(t)\}\{\tilde{W}_i(\theta,t) - \overline{W}(\theta,t)\}^{\otimes 2}\,dt,$$

$$\hat{A}_{\beta\gamma}(\theta) = n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)h\{\gamma^T X_i(t)\}\{\tilde{W}_i(\theta,t) - \overline{W}(\theta,t)\} \times\{\tilde{X}_i(\theta,t) - \overline{X}(\theta,t)\}^T\,d\hat{\Lambda}_0(\theta,t),$$

$$\hat{A}_{\gamma\beta}(\theta) = n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)h\{\gamma^T X_i(t)\}\{\tilde{X}_i(\theta,t) - \overline{X}(\theta,t)\} \times\{\tilde{W}_i(\theta,t) - \overline{W}(\theta,t)\}^T\,dt,$$

$$\hat{A}_{\gamma\gamma}(\theta) = n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)h\{\gamma^T X_i(t)\}\{\tilde{X}_i(\theta,t) - \overline{X}(\theta,t)\}^{\otimes 2}\,d\hat{\Lambda}_0(\theta,t).$$

If $Z = X$, then U_γ becomes the partial likelihood score function under the multiplicative hazard model; if $Z = W$ and $g(x) = x$, then U_β reduces to the ad hoc estimating function of Lin and Ying (1994) for model (1.2), which produces an explicit estimator.

To provide further insights into the conditions imposed in the previous developments, we consider the situation of i.i.d. $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$, $i = 1,\dots,n$. We make the following assumptions:

(A1) $\|Z_i(t)\| \le K$, a nonrandom constant;
(A2) g and h are continuously differentiable and $\{\partial D_i(\cdot,t)/\partial\theta,\ t\in[0,\tau],\ i\ge 1\}$ is equicontinuous in a neighborhood of θ_0;
(A3) $\sup_{0\le t\le\tau}\lambda_0(t) < \infty$ for every $\tau < \tau^*$, where $\tau^* = \inf\{t\colon EY_1(t) = 0\}$;
(A4) $\int_0^{\tau^*} EY_1(t)\{1 + \lambda_0(t)\}\,dt < \infty$.

Due to the law of large numbers, assumptions (A1)–(A3) ensure that the negligibility condition and stability conditions 1 and 2 required in Section 2.2 are satisfied for any $\tau < \tau^*$. Thus, under assumptions (A1)–(A3), we have the conclusion of Theorem 2.1 for any $\tau < \tau^*$. Furthermore, the conclusion of

Theorem 2.2 holds under an additional assumption on the nonsingularity of the slope matrix A. If assumption (A4) is also satisfied, then we may extend the convergence to $[0, \tau^*]$. The proof for this extension can be found in our technical report [Lin and Ying (1993)].

3. Estimation of $\Lambda_0(\cdot)$. Given $\hat{\theta}$, we may estimate $\Lambda_0(t)$ by

$$\hat{\Lambda}_0(\hat{\theta}, t) = \int_0^t \frac{\sum_{i=1}^n \left[dN_i(s) - Y_i(s) g\{\hat{\beta}^T W_i(s)\}\, ds\right]}{\sum_{i=1}^n Y_i(s) h\{\hat{\gamma}^T X_i(s)\}}. \tag{3.1}$$

It is convenient to introduce the notation

$$\begin{aligned}
a_g(t) &= \frac{\sum_{i=1}^n Y_i(t) g\{\beta_0^T W_i(t)\}}{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\}},\\
a_{g'}(t) &= \frac{\sum_{i=1}^n Y_i(t) g'\{\beta_0^T W_i(t)\} W_i(t)}{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\}},\\
a_{h'}(t) &= \frac{\sum_{i=1}^n Y_i(t) h'\{\gamma_0^T X_i(t)\} X_i(t)}{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\}},\\
a_0(t) &= n^{-1} \sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\},\\
a_{Dg}(t) &= \frac{\sum_{i=1}^n Y_i(t) g\{\beta_0^T W_i(t)\} D_i(\theta_0, t)}{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\}},\\
a_D(t) &= \frac{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\} D_i(\theta_0, t)}{\sum_{i=1}^n Y_i(t) h\{\gamma_0^T X_i(t)\}}.
\end{aligned} \tag{3.2}$$

THEOREM 3.1. *Suppose that all the assumptions in Theorem 2.2 are satisfied and that $a_g(t), \dots, a_D(t)$ defined by (3.2) converge in probability to $\tilde{a}_g(t), \dots, \tilde{a}_D(t)$ for every $t \le \tau$. Then $n^{1/2}\{\hat{\Lambda}_0(\hat{\theta}, \cdot) - \Lambda_0(\cdot)\}$ converges weakly in $\mathscr{D}[0, \tau]$ to a zero-mean Gaussian process with covariance function*

$$\begin{aligned}
\xi(t, s) = {}& \int_0^{t \wedge s} \frac{\{\tilde{a}_g(u) + \lambda_0(u)\}\, du}{\tilde{a}_0(u)} + \tilde{a}^T(t) A^{-1} V_\tau A^{-1^T} \tilde{a}(s)\\
& + \tilde{a}^T(t) A^{-1} \int_0^s \{\tilde{a}_{Dg}(u) - \tilde{a}_g(u) \tilde{a}_D(u)\}\, du\\
& + \tilde{a}^T(s) A^{-1} \int_0^t \{\tilde{a}_{Dg}(u) - \tilde{a}_g(u) \tilde{a}_D(u)\}\, du,
\end{aligned}$$

where

$$\tilde{a}(t) = -\begin{bmatrix} \int_0^t \tilde{a}_{g'}(u)\,du \\ \int_0^t \tilde{a}_{h'}(u)\lambda_0(u)\,du \end{bmatrix}.$$

PROOF. We decompose $n^{1/2}\{\hat{\Lambda}_0(\hat{\theta},t) - \Lambda_0(t)\}$ as

$$n^{1/2}\{\hat{\Lambda}_0(\hat{\theta},t) - \hat{\Lambda}_0(\theta_0,t)\} + n^{1/2}\{\hat{\Lambda}_0(\theta_0,t) - \Lambda_0^*(t)\} \quad (3.3)$$
$$+ n^{1/2}\{\Lambda_0^*(t) - \Lambda_0(t)\},$$

where

$$\Lambda_0^*(t) = \int_0^t 1_{[\sum_{i=1}^n Y_i(s)h\{\gamma_0^T X_i(s)\} > 0]}\lambda_0(s)\,ds.$$

The third term is asymptotically negligible. The second term is

$$n^{1/2}\{\hat{\Lambda}_0(\theta_0,t) - \Lambda_0^*(t)\} = n^{1/2}\int_0^t \frac{1_{[\sum_{i=1}^n Y_i(s)h\{\gamma_0^T X_i(s)\} > 0]}\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)h\{\gamma_0^T X_i(s)\}}. \quad (3.4)$$

By the Taylor expansion, the first term of (3.3) is

$$n^{1/2}\{\hat{\Lambda}_0(\hat{\theta},t) - \hat{\Lambda}_0(\theta_0,t)\} = \left\{\frac{\partial\hat{\Lambda}_0(\theta_0,t)}{\partial\theta}\right\}^T n^{1/2}(\hat{\theta} - \theta_0)$$
$$+ o_p(n^{1/2}\|\hat{\theta} - \theta_0\|),$$

where o_p is uniform in $t \le \tau$. Note that

$$\frac{\partial\hat{\Lambda}_0(\theta_0,t)}{\partial\theta} = -\begin{bmatrix} \int_0^t a_{g'}(s)\,ds \\ \int_0^t a_{h'}(s)\lambda_0(s)\,ds + n^{-1}\sum_{i=1}^n \int_0^t \left\{\frac{a_{h'}(s)}{a_0(s)}\right\} dM_i(s) \end{bmatrix}.$$

By the triangle inequality,

$$\sup_{0\le t\le\tau}\left\|\frac{\partial\hat{\Lambda}_0(\theta_0,t)}{\partial\theta} - \tilde{a}(t)\right\|$$
$$\le \int_0^\tau \{\|a_{g'}(t) - \tilde{a}_{g'}(t)\| + \|a_{h'}(t) - \tilde{a}_{h'}(t)\|\lambda_0(t)\}\,dt \quad (3.5)$$
$$+ n^{-1}\sup_{0\le t\le\tau}\left\|\int_0^t \left\{\frac{a_{h'}(s)}{a_0(s)}\right\}\sum_{i=1}^n dM_i(s)\right\|.$$

The first term on the right-hand side of inequality (3.5) is $o_p(1)$ by the convergence of $a_{g'}$ and $a_{h'}$ to $\tilde{a}_{g'}$ and $\tilde{a}_{h'}$. The second term is also $o_p(1)$ by Lenglart's inequality [Andersen, Borgan, Gill and Keiding (1993), page 86]. It follows from these results and the asymptotic expression for $n^{1/2}(\hat{\theta} - \theta_0)$ derived in the proof of Theorem 2.2 that, uniformly in t,

$$n^{1/2}\{\hat{\Lambda}_0(\hat{\theta}, t) - \hat{\Lambda}_0(\theta_0, t)\} = \tilde{a}^T(t) A^{-1} n^{-1/2} S(\theta_0, \tau) + o_p(1),$$

which together with (3.4) implies that, uniformly in t,

$$\begin{aligned} &n^{1/2}\{\hat{\Lambda}_0(\hat{\theta}, t) - \Lambda_0(t)\} \\ &= n^{1/2} \int_0^t \frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s) h\{\gamma_0^T X_i(s)\}} \\ &\quad + \tilde{a}^T(t) A^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \{D_i(\theta_0, s) - \overline{D}(\theta_0, s)\}\, dM_i(s) + o_p(1). \end{aligned} \tag{3.6}$$

As in the proof of Theorem 2.1, the desired weak convergence now follows from Rebolledo's central limit theorem together with a straightforward covariance calculation. □

REMARK 3.1. In the i.i.d. case discussed in Section 2.4, the conclusion of Theorem 3.1 holds for any $\tau < \tau^*$ provided that A is nonsingular. This is because the a's in (3.2) converge to the $\tilde{a}$'s by the law of large numbers.

It is natural to estimate the covariance function $\xi(t, s)$ by

$$\begin{aligned} \hat{\xi}(t, s) &= \int_0^{t \wedge s} \frac{a_g(\hat{\theta}, u)\, du + d\hat{\Lambda}_0(\hat{\theta}, u)}{a_0(\hat{\theta}, u)} \\ &\quad + a^T(\hat{\theta}, t) \hat{A}^{-1}(\hat{\theta}) \hat{V}(\hat{\theta}) \hat{A}^{-1}(\hat{\theta})^T a(\hat{\theta}, s) \\ &\quad + a^T(\hat{\theta}, t) \hat{A}^{-1}(\hat{\theta}) \int_0^s \{a_{Dg}(\hat{\theta}, u) - a_g(\hat{\theta}, u) a_D(\hat{\theta}, u)\}\, du \\ &\quad + a^T(\hat{\theta}, s) \hat{A}^{-1}(\hat{\theta}) \int_0^t \{a_{Dg}(\hat{\theta}, u) - a_g(\hat{\theta}, u) a_D(\hat{\theta}, u)\}\, du, \end{aligned}$$

where

$$a(\hat{\theta}, t) = -\begin{bmatrix} \int_0^t a_{g'}(\hat{\theta}, u)\, du \\ \int_0^t a_{h'}(\hat{\theta}, u)\, d\hat{\Lambda}_0(\hat{\theta}, u) \end{bmatrix},$$

and $a_g(\hat{\theta}, t), \dots, a_D(\hat{\theta}, t)$ are obtained from $a_g(t), \dots, a_D(t)$ defined in (3.2) by substituting $\hat{\theta}$ for θ_0. Note that $\hat{A}$ and $\hat{V}$ were defined in Section 2.3. The following corollary establishes the consistency of $\hat{\xi}$ as well as that of $\hat{A}^{-1}(\hat{\theta})\hat{V}(\hat{\theta})\hat{A}^{-1}(\hat{\theta})^T$ proposed in Section 2.3 to estimate $A^{-1}V_\tau A^{-1^T}$.

COROLLARY 3.2. *Suppose that the assumptions of Theorem* 3.1 *are satisfied and that the total variations of $D_i(\theta_0, t)$ and $Z_i(t)$ are bounded uniformly over all i. Then $\hat{A}^{-1}(\hat{\theta})\hat{V}(\hat{\theta})\hat{A}^{-1}(\hat{\theta})^T \to_P A^{-1}V_\tau A^{-1^T}$ and $\hat{\xi}(t, s) \to_P \xi(t, s)$.*

PROOF. Since $\hat{\theta} \to_P \theta_0$, the smoothness of D_i, g and h implies that

$$\hat{A}(\hat{\theta}) = \int_0^\tau n^{-1} \sum_{i=1}^n Y_i(t)\{D_i(\theta_0, t) - \bar{D}(\theta_0, t)\}$$
$$\times \left[g'\{\beta_0^T W_i(t)\}W_i^T(t)\,dt, h'\{\gamma_0^T X_i(t)\}X_i^T(t)\,d\hat{\Lambda}_0(\hat{\theta}, t)\right] + o_p(1).$$

Furthermore, the integrand in

$$\int_0^\tau n^{-1} \sum_{i=1}^n Y_i(t)\{D_i(\theta_0, t) - \bar{D}(\theta_0, t)\}h'\{\gamma_0^T X_i(t)\}X_i^T(t)\,d\hat{\Lambda}_0(\hat{\theta}, t)$$

is of bounded variation and, by Theorem 3.1, $\hat{\Lambda}_0(\hat{\theta}, t)$ converges to $\Lambda_0(t)$ uniformly. Thus integration by parts yields $\hat{A}(\hat{\theta}) = -n^{-1}\,\partial S(\theta_0, \tau)/\partial\theta + o_p(1)$, which of course converges to A. Such arguments, together with a simple use of Lenglart's inequality, show that $\hat{V}(\hat{\theta})$ converges to V_τ. Likewise, the consistency of $\hat{\xi}(t, s)$ can be verified. The details are omitted. □

REMARK 3.2. Estimator (3.1) may not always be nondecreasing in t. A simple modification which ensures the monotonicity is $\hat{\Lambda}_0^*(\theta, t) = \sup_{0 \le s \le t} \hat{\Lambda}_0(\theta, s)$. It can be shown that $n^{1/2}\{\hat{\Lambda}_0^*(\hat{\theta}, \cdot) - \Lambda_0(\cdot)\}$ converges weakly to the same limiting Gaussian process as $n^{1/2}\{\hat{\Lambda}_0(\hat{\theta}, \cdot) - \Lambda_0(\cdot)\}$ [Lin and Ying (1993)]. In the special setting where the counting process takes at most one jump, one may construct approximate confidence bands for $\Lambda_0(\cdot)$ by using a Monte Carlo technique described in Lin, Fleming and Wei (1994).

4. Numerical results. In this section, we present some results from our numerical studies on the methods developed in the previous two sections.

4.1. *Simulation studies.* Monte Carlo experiments were conducted to assess the adequacy of the proposed large-sample approximations for practical sample sizes. Table 1 displays some typical results. For this table, failure times were generated from the additive–multiplicative model $\lambda(t \mid W, X) = \beta_0 W + \lambda_0(t)\exp(\gamma_0 X)$ and censoring times from the uniform $(0, \kappa)$ distribution, where W and X are independent uniform $(0, 1)$ variables, λ_0 has a Weibull distribution and κ is chosen to ensure a desired censoring probability; the estimating function $[U_\beta, U_\gamma]$ described in Section 2.4 was used. It is evident that the estimators are nearly unbiased and the associated tests and confidence intervals have proper sizes and coverage probabilities provided that the number of uncensored failure times is not too small.

4.2. *A real example.* We now illustrate the proposed methods with the lung cancer data presented by Ying, Jung and Wei (1995). One hundred twenty-three patients with small cell lung cancer were randomly assigned to two treatment regimens, the first of which administered cisplatin followed by etoposide and the second of which administered etoposide followed by cisplatin. By the end of the study, 49 of the 64 patients on treatment 1 and 51 of the 59 patients on treatment 2 had died. The investigators were interested in relating the survival time to the treatment variable Z_1 (indicating, by the value 0 versus 1, whether the patient was on treatment 1 or 2) and the age at

TABLE 1

Monte Carlo estimates for the sampling mean and variance of $\hat{\beta}$, for the sampling mean of the variance estimator $\widehat{\mathrm{var}}(\hat{\beta})$ *and for the size of the* 0.05*-level Wald test for testing* H_0: $\beta = \beta_0$ *under the model* $\lambda(t \mid W, X) = \beta_0 W + \lambda_0(t)\exp(0.5X)$

			$\beta_0 = 0$				$\beta_0 = 0.5$			
$\lambda_0(t)$	n	Censoring probability	Mean of $\hat{\beta}$	Variance of $\hat{\beta}$	Mean of $\widehat{\mathrm{var}}(\hat{\beta})$	Size of test	Mean of $\hat{\beta}$	Variance of $\hat{\beta}$	Mean of $\widehat{\mathrm{var}}(\hat{\beta})$	Size of test
1	50	20%	0.001	0.647	0.628	0.045	0.519	0.950	0.917	0.049
1	50	50%	−0.009	0.980	0.983	0.046	0.484	1.444	1.418	0.056
1	100	20%	0.001	0.281	0.271	0.048	0.512	0.405	0.395	0.044
1	100	50%	0.003	0.451	0.438	0.048	0.507	0.654	0.636	0.048
1	100	70%	0.002	0.719	0.700	0.047	0.486	1.018	1.020	0.050
$2t$	50	20%	−0.005	0.460	0.466	0.048	0.504	0.596	0.613	0.050
$2t$	50	50%	−0.018	0.470	0.473	0.054	0.469	0.623	0.643	0.060
$2t$	100	20%	0.001	0.218	0.214	0.047	0.509	0.283	0.285	0.046
$2t$	100	50%	−0.003	0.224	0.225	0.048	0.496	0.299	0.311	0.046
$2t$	100	70%	−0.001	0.223	0.221	0.046	0.484	0.311	0.324	0.054

Note: *Each entry is based on* 10,000 *replications.*

TABLE 2
Regression analyses of the lung cancer data

Parameters	Cox model	Additive–multiplicative model
Treatment		
Est.	0.463	0.00052
S.E.	0.202	0.00026
Est./S.E.	2.29	2.02
Age		
Est.	0.0278	0.0350
S.E.	0.0127	0.0174
Est./S.E.	2.18	2.01

study entry Z_2. The left side of Table 2 shows the results for the Cox model $\lambda(t \mid Z) = \lambda_0(t)\exp(\gamma_{01}Z_1 + \gamma_{02}Z_2)$. Goodness-of-fit analysis revealed that the proportional hazards assumption might not be completely satisfied with respect to Z_1: the hazard ratio appears to be larger early on and smaller toward the end. In general, if $\lambda_0(\cdot)$ is increasing over time (as is often the case with chronic diseases) and the hazard ratio is decreasing over time, then it may be more appropriate to postulate an additive rather than a multiplicative covariate effect. Thus, we also fit the additive–multiplicative model $\lambda(t \mid Z) = \beta_0 Z_1 + \lambda_0(t)\exp(\gamma_0 Z_2)$; the right side of Table 2 displays the results based on $[U_\beta, U_\gamma]$ of Section 2.4. As a rough check on the goodness of fit, Figure 1 compares the model-based estimates of survival probabilities with the local Kaplan–Meier estimates. (For the graphical analysis, we dichotomized Z_2 at its median value 63.) The figure suggests that both models provide fairly good summarization of the data, but the models seem to underestimate the treatment difference for $t < 600$. A plot of the cumulative hazard function estimates (not shown here) indicated that $\lambda_0(t)$ is roughly constant, although slightly decreasing for small values of t and slightly increasing for large values of t. In short, treatment 1 is more beneficial than treatment 2, and younger patients have better prognosis than older ones; the treatment difference appears to be a bit larger in the early stage of therapy than in the late stage, with an average risk ratio of about 1.6 and an average risk difference of about 0.0005.

5. Asymptotic efficiency. Sections 2 and 3 provided some simple estimators for θ_0 and Λ_0. In this section, we shall assess the asymptotic efficiency of those estimators and construct estimators that are asymptotically efficient. To avoid introducing stability conditions, we assume that $\{N_i(t), Y_i(t), Z_i(t), t \in [0, \tau]\}$, $i = 1, \dots, n$, are i.i.d. In addition to assumptions (A1) and (A2) made in Section 2.4, we further assume that the following hold: (i) λ_0 is continuous; (ii) there exists a constant $\varepsilon_\tau > 0$ such that, with probability 1,

$$\inf_{0 \le t \le \tau} \min\left[h\{\gamma_0^T X_1(t)\}, g\{\beta_0^T W_1(t)\} + h\{\gamma_0^T X_1(t)\}\lambda_0(t)\right] \ge \varepsilon_\tau;$$

(iii) $E\{\inf_{0 \le t \le \tau} Y_1(t)\} > 0$.

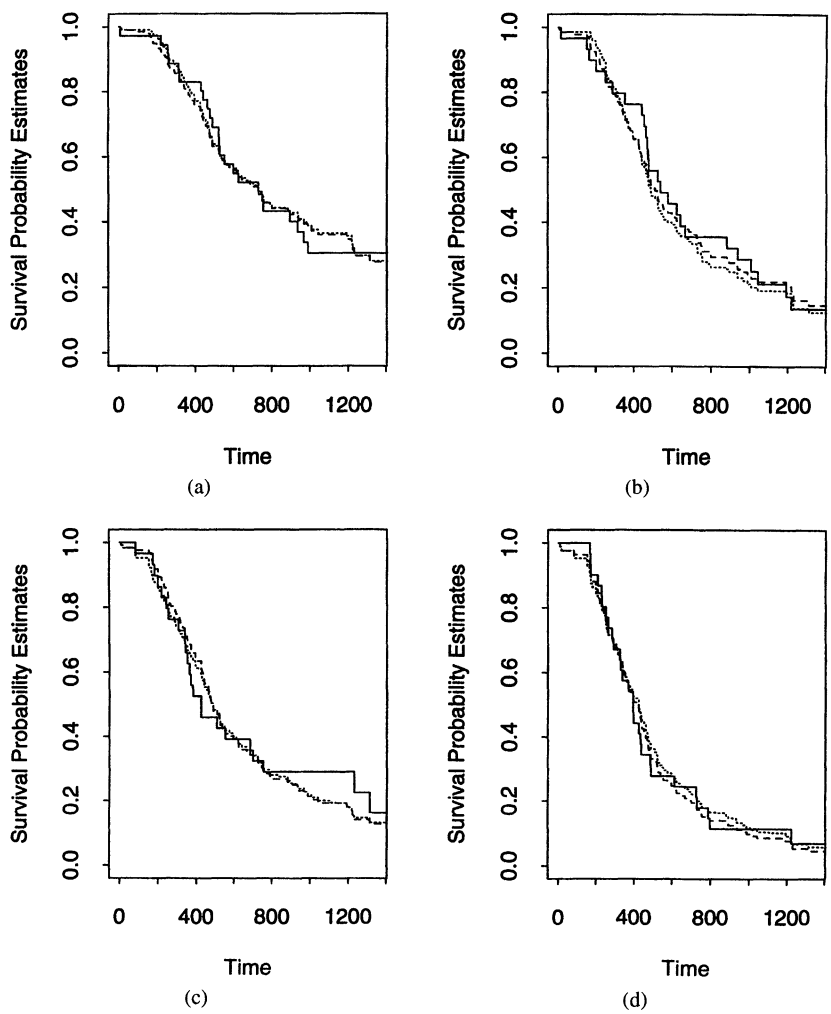

FIG. 1. *Survival probability estimates based on the Kaplan–Meier method, the Cox model and the additive–multiplicative model, shown by the solid, dashed and dotted curves, respectively, for four risk groups*: (a) *treatment* 1, (*age*) ≤ 63; (b) *treatment* 1, (*age*) > 63; (c) *treatment* 2, (*age*) ≤ 63; (d) *treatment* 2, (*age*) > 63.

5.1. *Efficient estimation of* θ_0. It is convenient to introduce the following notation:

$$\tilde{W}_i(\theta;t) = \frac{g'\{\beta^T W_i(t)\}W_i(t)}{h\{\gamma^T X_i(t)\}}, \qquad \tilde{W}_i(t) = \tilde{W}_i(\theta_0;t);$$

$$\tilde{X}_i(\theta;t) = \frac{h'\{\gamma^T X_i(t)\}X_i(t)}{h\{\gamma^T X_i(t)\}}, \qquad \tilde{X}_i(t) = \tilde{X}_i(\theta_0;t);$$

$$\mu_i(\theta, \lambda; t) = \frac{h^2\{\gamma^T X_i(t)\}}{g\{\beta^T W_i(t)\} + h\{\gamma^T X_i(t)\}\lambda(t)}, \qquad \mu_i(t) = \mu_i(\theta_0, \lambda_0; t);$$

$$\overline{W}_\mu(\theta, \lambda; t) = \frac{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t) \tilde{W}_i(\theta; t)}{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t)}, \qquad \overline{w}_\mu(t) = \lim_{n\to\infty} \overline{W}_\mu(\theta_0, \lambda_0; t);$$

$$\overline{X}_\mu(\theta, \lambda; t) = \frac{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t) \tilde{X}_i(\theta; t)}{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t)}, \qquad \overline{x}_\mu(t) = \lim_{n\to\infty} \overline{X}_\mu(\theta_0, \lambda_0; t);$$

$$\tilde{Z}_i(\theta, \lambda; t) = \begin{bmatrix} \tilde{W}_i(\theta; t) \\ \tilde{X}_i(\theta; t)\lambda(t) \end{bmatrix}, \qquad \tilde{Z}_i(t) = \tilde{Z}_i(\theta_0, \lambda_0; t);$$

$$\overline{Z}_\mu(\theta, \lambda; t) = \frac{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t) \tilde{Z}_i(\theta, \lambda; t)}{\sum_{i=1}^n \mu_i(\theta, \lambda; t) Y_i(t)}, \qquad \overline{z}_\mu(t) = \lim_{n\to\infty} \overline{Z}_\mu(\theta_0, \lambda_0; t).$$

Due to assumptions (ii) and (iii), the denominators in the above expressions are nonzero for all $t \in [0, \tau]$, θ sufficiently close to θ_0 and n sufficiently large. Hence, all the quantities are well defined.

To derive the information bound for estimating θ_0, consider the following parametric submodel:

$$\lambda(t \mid Z; \theta, \eta) = g\{\beta^T W(t)\} + h\{\gamma^T X(t)\}\Big[\lambda_0(t)\big\{1 + \eta_2^T \overline{x}_\mu(t)\big\} + \eta_1^T \overline{w}_\mu(t)\Big],$$

where $\eta = (\eta_1^T, \eta_2^T)^T$. Note that, in this model, θ and η are unknown parameters whereas $\lambda_0(\cdot)$ and $\{\overline{x}_\mu(\cdot), \overline{w}_\mu(\cdot)\}$ are fixed functions. The ($2p$-dimensional) likelihood score function for $(\theta^T, \eta^T)^T$ evaluated at $\theta = \theta_0$ and $\eta = 0$ is

$$\sum_{i=1}^n \int_0^\tau \frac{\tilde{Z}_i(t)\mu_i(t)}{h\{\gamma_0^T X_i(t)\}}\, dM_i(t),$$
$$\sum_{i=1}^n \int_0^\tau \frac{\overline{z}_\mu(t)\mu_i(t)}{h\{\gamma_0^T X_i(t)\}}\, dM_i(t).$$

The corresponding Fisher information matrix (normalized by n) is

$$\begin{bmatrix} I_{\theta\theta} & I_{\theta\eta} \\ I_{\eta\theta} & I_{\eta\eta} \end{bmatrix} = E\Bigg(n^{-1} \sum_{i=1}^n \int_0^\tau \begin{bmatrix} \tilde{Z}_i(t) \\ \overline{z}_\mu(t) \end{bmatrix}^{\otimes 2} \frac{\mu_i^2(t)}{h^2\{\gamma_0^T X_i(t)\}} Y_i(t)$$
$$\times \big[g\{\beta_0^T W_i(t)\} + h\{\gamma_0^T X_i(t)\}\lambda_0(t)\big]\, dt\Bigg)$$
$$= E\int_0^\tau \begin{bmatrix} \tilde{Z}_1(t) \\ \overline{z}_\mu(t) \end{bmatrix}^{\otimes 2} Y_1(t)\mu_1(t)\, dt.$$

Thus, the information for θ_0 is

$$I_0 = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta} = E\int_0^\tau \left\{\tilde{Z}_1(t) - \bar{z}_\mu(t)\right\}^{\otimes 2} Y_1(t)\mu_1(t)\, dt,$$

and I_0^{-1} is the information bound for estimating θ_0 in the above parametric submodel [Bickel, Klaassen, Ritov and Wellner (1993), page 28]. By definition, the information bound for an infinite-dimensional model is the supremum of the information bounds among all regular parametric submodels [Bickel, Klaassen, Ritov and Wellner (1993), page 46]. Therefore, if we can construct a regular semiparametric estimator whose asymptotic covariance matrix (normalized by n) is I_0^{-1}, then we can claim that I_0^{-1} is the (semiparametric) information bound for estimating θ_0 when $\lambda_0(\cdot)$ is completely unspecified and that the estimator so constructed is asymptotically efficient [Bickel, Klaassen, Ritov and Wellner (1993), pages 76–77].

To ease our theoretical analysis, we shall employ the so-called sample-splitting technique to construct an asymptotically efficient estimator for θ_0. We randomly partition the entire sample into two groups, the first one containing (the first) $[n/2]$ (the largest integer less than or equal to $n/2$) subjects. Let $\hat{\theta}^{(1)}$ denote an initial $n^{1/2}$-consistent estimator of θ_0 on the basis of the first group of the data by using an inefficient estimating equation given in Section 2. We can then apply the kernel smoothing method of Ramlau-Hansen (1983) to the estimator of Λ_0 given in Section 3 to obtain a consistent estimator $\hat{\lambda}_0^{(1)}$ of λ_0. Note that to guarantee consistency we can choose bandwidth to be $n^{-1/3}$ and make constant extrapolation on $[0, 2n^{-1/3}]$ and $[\tau - 2n^{-1/3}, \tau]$. Likewise, we can construct consistent estimators $\hat{\theta}^{(2)}$ and $\hat{\lambda}_0^{(2)}$ from the second group of the data.

Given the initial consistent estimators $\hat{\theta}^{(k)}$ and $\hat{\lambda}_0^{(k)}$, $k = 1, 2$, we define

$$\begin{aligned} S_{\mathrm{opt}}(\theta) = &\sum_{i=1}^{[n/2]} \int_0^\tau \left\{\tilde{Z}_i\left(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\right) - \bar{Z}_\mu^{(1)}\left(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\right)\right\} \\ &\times \frac{\mu_i\left(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\right)}{h\{\gamma^T X_i(t)\}} \left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right] \\ &+ \sum_{i=[n/2]+1}^{n} \int_0^\tau \left\{\tilde{Z}_i\left(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\right) - \bar{Z}_\mu^{(2)}\left(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\right)\right\} \\ &\times \frac{\mu_i\left(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\right)}{h\{\gamma^T X_i(t)\}} \left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right], \end{aligned}$$

where

$$\bar{Z}_\mu^{(1)}(\theta, \lambda; t) = \frac{\sum_{i=1}^{[n/2]} \mu_i(\theta, \lambda; t) Y_i(t) \tilde{Z}_i(\theta, \lambda; t)}{\sum_{i=1}^{[n/2]} \mu_i(\theta, \lambda; t) Y_i(t)},$$

$$\bar{Z}_\mu^{(2)}(\theta, \lambda; t) = \frac{\sum_{i=[n/2]+1}^{n} \mu_i(\theta, \lambda; t) Y_i(t) \tilde{Z}_i(\theta, \lambda; t)}{\sum_{i=[n/2]+1}^{n} \mu_i(\theta, \lambda; t) Y_i(t)}.$$

A simple algebraic manipulation yields

$$
\begin{aligned}
S_{\mathrm{opt}}(\theta) = \sum_{i=1}^{[n/2]} \int_0^{\tau} &\Big\{\tilde{Z}_i\big(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\big) - \bar{Z}_{\mu}^{(1)}\big(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\big)\Big\} \\
&\times \frac{\mu_i\big(\hat{\beta}^{(2)}, \gamma, \hat{\lambda}_0^{(2)}; t\big)}{h\{\gamma^T X_i(t)\}}\, dM_i(\theta, t) \\
+ \sum_{i=[n/2]+1}^{n} \int_0^{\tau} &\Big\{\tilde{Z}_i\big(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\big) - \bar{Z}_{\mu}^{(2)}\big(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\big)\Big\} \\
&\times \frac{\mu_i\big(\hat{\beta}^{(1)}, \gamma, \hat{\lambda}_0^{(1)}; t\big)}{h\{\gamma^T X_i(t)\}}\, dM_i(\theta, t).
\end{aligned} \tag{5.1}
$$

From (5.1) we can see heuristically that S_{opt} should result in an asymptotically efficient estimator, that is, an estimator whose asymptotic covariance matrix is approximately $(nI_0)^{-1}$. The following theorem formalizes this heuristic.

THEOREM 5.1. *Suppose that the assumptions made at the beginning of this section are satisfied and that the information matrix I_0 is nonsingular. Then there exists a neighborhood of θ_0 within which, with probability tending to 1 as $n \to \infty$, the root $\hat{\theta}_{\mathrm{opt}}$ of $S_{\mathrm{opt}}(\theta) = 0$ is uniquely defined and $n^{1/2}(\hat{\theta}_{\mathrm{opt}} - \theta_0) \to_L \mathcal{N}(0, I_0^{-1})$.*

PROOF. Let

$$
R_i(\theta, \lambda; t) = \Big\{\tilde{Z}_i(\theta, \lambda; t) - \bar{Z}_{\mu}^{(1)}(\theta, \lambda; t)\Big\} \frac{\mu_i(\theta, \lambda; t)}{h\{\gamma^T X_i(t)\}}, \qquad i = 1, \dots, n.
$$

Then the derivative at $\theta = \theta_0$ for the first summation on the right-hand side of (5.1) is

$$
\begin{aligned}
&\sum_{i=1}^{[n/2]} \int_0^{\tau} \frac{\partial R_i\big(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\big)}{\partial \gamma}\, dM_i(t) \\
&\quad + \sum_{i=1}^{[n/2]} \int_0^{\tau} R_i\big(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\big)\, d\frac{\partial M_i(\theta_0, t)}{\partial \theta}.
\end{aligned}
$$

The first term is $O_p(n^{1/2})$ because it is a sum of integrals of predictable processes with respect to the martingales $\{M_i(t); i = 1, \dots, [n/2]\}$, where the σ-filtration $\mathscr{F}_t^{(1)}$ is generated by

$$
\Big\{N_i(s), Y_i(s+), Z_i(s+); s \le t, 1 \le i \le \Big[\frac{n}{2}\Big]\Big\}
$$

and

$$
\Big\{N_j(u), Y_j(u), Z_j(u); 0 \le u \le \tau, \Big[\frac{n}{2}\Big] < j \le n\Big\}.
$$

Since $\hat{\beta}^{(2)}$ and $\hat{\lambda}_0^{(2)}$ are consistent, the second term is approximately

$$\sum_{i=1}^{[n/2]} \int_0^\tau R_i(\theta_0, \lambda_0; t)\, d\frac{\partial M_i(\theta_0, t)}{\partial \theta},$$

which is easily seen to be $-(n/2)I_0 + o_p(n)$ by the law of large numbers. Thus

$$n^{-1}\left\{ \sum_{i=1}^{[n/2]} \int_0^\tau \frac{\partial R_i\left(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\right)}{\partial \gamma}\, dM_i(t) + \sum_{i=1}^{[n/2]} \int_0^\tau R_i\left(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\right) d\frac{\partial M_i(\theta_0, t)}{\partial \theta} \right\} \to_P -\frac{I_0}{2}.$$

With a similar result for the second sum in (5.1) we get $n^{-1}\, \partial S_{\mathrm{opt}}(\theta_0)/\partial\theta \to_P -I_0$. Therefore, by the arguments given in the proof of Theorem 2.2, there exists a neighborhood of θ_0 within which, with probability tending to 1 as $n \to \infty$, $\hat{\theta}_{\mathrm{opt}}$ is uniquely defined and converges to θ_0. By the Taylor expansion of $S_{\mathrm{opt}}(\hat{\theta}_{\mathrm{opt}})$ at θ_0, we get

$$
\begin{aligned}
n^{1/2}\left(\hat{\theta}_{\mathrm{opt}} - \theta_0\right) &= \left\{I_0^{-1} + o_p(1)\right\} n^{-1/2} S_{\mathrm{opt}}(\theta_0) \\
&= I_0^{-1} n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \left\{\tilde{Z}_i(t) - \bar{z}_\mu(t)\right\} \frac{\mu_i(t)}{h\{\gamma_0^T X_i(t)\}}\, dM_i(t) \\
&\quad + o_p(1),
\end{aligned}
\tag{5.2}
$$

where the second equality comes from

$$n^{-1/2} \sum_{i=1}^{[n/2]} \int_0^\tau \Bigg[\left\{\tilde{Z}_i\left(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\right) - \bar{Z}_\mu^{(1)}\left(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\right)\right\} \frac{\mu_i\left(\hat{\beta}^{(2)}, \gamma_0, \hat{\lambda}_0^{(2)}; t\right)}{h\{\gamma_0^T X_i(t)\}} - \left\{\tilde{Z}_i(t) - \bar{z}_\mu(t)\right\} \frac{\mu_i(t)}{h\{\gamma_0^T X_i(t)\}} \Bigg] dM_i(t) = o_p(1)$$

and a similar approximation for the second sum in (5.1). It follows from (5.2) that $n^{1/2}(\hat{\theta} - \theta_0) \to_L \mathcal{N}(0, I_0^{-1})$. □

Remark 5.1. The ad hoc estimating function suggested at the beginning of Section 2.4 will be close to S_{opt} if λ_0 is roughly constant and g is small relative to $h\lambda_0$. By the arguments given in the proof of Corollary 3.2, I_0 can be consistently estimated by $\hat{V}_{\mathrm{opt}}(\hat{\theta}_{\mathrm{opt}})$, where $\hat{V}_{\mathrm{opt}}(\theta)$ is obtained from $\hat{V}(\theta)$ defined at the end of Section 2.3 with obvious changes of the integrands. For moderate-sized samples, it is more accurate to estimate the limiting covariance matrix of $n^{1/2}(\hat{\theta}_{\mathrm{opt}} - \theta_0)$ by $\hat{A}_{\mathrm{opt}}^{-1}(\hat{\theta}_{\mathrm{opt}})\hat{V}_{\mathrm{opt}}(\hat{\theta}_{\mathrm{opt}})\hat{A}_{\mathrm{opt}}^{-1}(\hat{\theta}_{\mathrm{opt}})^T$ than by $\hat{V}_{\mathrm{opt}}^{-1}(\hat{\theta}_{\mathrm{opt}})$, where $\hat{A}_{\mathrm{opt}}(\theta)$ is the obvious modification of $\hat{A}(\theta)$.

5.2. *Efficient estimation of* Λ_0. In this section, we discuss the efficient estimation of $\alpha_0 \equiv \Lambda_0(t_0)$ for a given time point t_0 along the lines of

Andersen, Borgan, Gill and Keiding [(1993), VIII.2.4 and VIII.4.3]. To motivate our procedure, suppose for the moment that the regression parameters β_0 and γ_0 are known. Consider the following class of estimators for α_0:

$$\hat{\Lambda}_c(\theta_0, t_0) = \int_0^{t_0} \frac{\sum_{i=1}^n c_i(t)\left[dN_i(t) - Y_i(t) g\{\beta_0^T W_i(t)\}\, dt\right]}{\sum_{i=1}^n c_i(t) h\{\gamma_0^T X_i(t)\} Y_i(t)},$$

where $\{c_i(\cdot)\}$ are positive predictable processes.

It is easy to see that $\hat{\Lambda}_c(\theta_0, t_0)$ is asymptotically unbiased with approximate variance

$$\int_0^{t_0} \frac{\sum_{i=1}^n c_i^2(t)\left[g\{\beta_0^T W_i(t)\} + h\{\gamma_0^T X_i(t)\}\lambda_0(t)\right] Y_i(t)\, dt}{\left[\sum_{i=1}^n c_i(t) h\{\gamma_0^T X_i(t)\} Y_i(t)\right]^2}$$
$$\geq \int_0^{t_0} \frac{dt}{\sum_{i=1}^n \mu_i(t) Y_i(t)},$$

where the equality holds if $c_i(t) = \mu_i(t)/h\{\gamma_0^T X_i(t)\}$. Writing $y_\mu(t) = E\{\mu_1(t) Y_1(t)\}$, we next show that $\int_0^{t_0} y_\mu^{-1}(t)\, dt$ is actually a lower bound for the variance (normalized by n) of any regular estimator of $\Lambda_0(t_0)$ (when θ_0 is known). To this end, we introduce a one-parameter submodel

$$\lambda_0(\alpha; t) = \lambda_0(t) + (\alpha - \alpha_0)\tilde{\mu}_0(t),$$

where $\tilde{\mu}_0(t) = 1_{(t \leq t_0)}/\{y_\mu(t) \int_0^{t_0} y_\mu^{-1}(s)\, ds\}$. Clearly, $\Lambda_0(\alpha; t_0) = \alpha$. It is straightforward to show that the Fisher information (again normalized by n) for α at α_0 is $\{\int_0^{t_0} y_\mu^{-1}(t)\, dt\}^{-1}$.

Since the unknown θ_0 can be efficiently estimated by $\hat{\theta}_{\text{opt}}$, it is natural to speculate that

$$\hat{\Lambda}_\omega\left(\hat{\theta}_{\text{opt}}, t_0\right) = \int_0^{t_0} \frac{\sum_{i=1}^n \omega_i(t)\left[dN_i(t) - Y_i(t) g\left\{\hat{\beta}_{\text{opt}}^T W_i(t)\right\} dt\right]}{\sum_{i=1}^n \omega_i(t) h\left\{\hat{\gamma}_{\text{opt}}^T X_i(t)\right\} Y_i(t)},$$

with $\omega_i(t) = \mu_i(t)/h\{\gamma_0^T X_i(t)\}$, should be asymptotically optimal. Analogously to (3.6),

$$n^{1/2}\left\{\hat{\Lambda}_\omega\left(\hat{\theta}_{\text{opt}}, t_0\right) - \Lambda_0(t_0)\right\}$$
$$= n^{1/2} \int_0^{t_0} \frac{\sum \omega_i(t)\, dM_i(t)}{\sum \omega_i(t) h\{\gamma_0^T X_i(t)\} Y_i(t)}$$
$$- \int_0^{t_0} \bar{z}_\mu^T(t)\, dt\, I_0^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \left\{\tilde{Z}_i(t) - \bar{Z}_\mu(t)\right\}\omega_i(t)\, dM_i(t) + o_p(1),$$

the limiting variance being $v_0 = \int_0^{t_0} y_\mu^{-1}(t)\, dt + \int_0^{t_0} \bar{z}_\mu^T(t)\, dt\, I_0^{-1} \int_0^{t_0} \bar{z}_\mu(t)\, dt$.

We claim that v_0 is the optimal variance for the estimation of $\Lambda_0(t_0)$. We shall establish this claim by providing a one-parameter family whose Fisher information is v_0^{-1}. Define

$$\lambda_0(\alpha;t) = \lambda_0(t) + (\alpha - \alpha_0)\left\{1_{(t \le t_0)} y_\mu^{-1}(t) + \int_0^{t_0} \bar{z}_\mu^T(s)\, ds\, I_0^{-1} \bar{z}_\mu(t)\right\} v_0^{-1},$$

$$\theta = \theta_0 - (\alpha - \alpha_0) I_0^{-1} \int_0^{t_0} \bar{z}_\mu(s)\, ds\, v_0^{-1}.$$

Obviously, $\alpha = \int_0^{t_0} \lambda_0(\alpha; t)\, dt$. The likelihood score function at α_0 can be shown to be

$$v_0^{-1}\Bigg[\sum_{i=1}^n \int_0^\tau 1_{(t \le t_0)} y_\mu^{-1}(t) \frac{\mu_i(t)}{h\{\gamma_0^T X_i(t)\}}\, dM_i(t)$$
$$- \int_0^{t_0} \bar{z}_\mu^T(t)\, dt\, I_0^{-1} \sum_{i=1}^n \int_0^\tau \left\{\tilde{Z}_i(t) - \bar{z}_\mu(t)\right\} \frac{\mu_i(t)}{h\{\gamma_0^T X_i(t)\}}\, dM_i(t)\Bigg].$$

Thus the normalized Fisher information converges to v_0^{-1}.

We may approximate the unknown optimal weights $\omega_i(t) = \mu_i(t)/h\{\gamma_0^T X_i(t)\}$ by applying the sample-splitting technique again. Let $\omega_i(\theta, \lambda; t) = \mu_i(\theta, \lambda; t)/h\{\gamma^T X_i(t)\}$, and let $\hat{\theta}^{(k)}$, $\hat{\lambda}_0^{(k)}$, $k = 1, 2$, and $\hat{\theta}_{\text{opt}}$ be as defined in Section 5.1. In addition, define

$$\hat{\Lambda}_{\text{opt}}(\theta, t_0) = \sum_{i=1}^{[n/2]} \int_0^{t_0} \frac{\omega_i\left(\hat{\theta}^{(2)}, \hat{\lambda}_0^{(2)}; t\right)\left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right]}{2\sum_{i=1}^{[n/2]} \omega_i\left(\hat{\theta}^{(2)}, \hat{\lambda}_0^{(2)}; t\right) h\{\gamma^T X_i(t)\} Y_i(t)}$$
$$+ \sum_{i=[n/2]+1}^{n} \int_0^{t_0} \frac{\omega_i\left(\hat{\theta}^{(1)}, \hat{\lambda}_0^{(1)}; t\right)\left[dN_i(t) - Y_i(t) g\{\beta^T W_i(t)\}\, dt\right]}{2\sum_{i=[n/2]+1}^{n} \omega_i\left(\hat{\theta}^{(1)}, \hat{\lambda}_0^{(1)}; t\right) h\{\gamma^T X_i(t)\} Y_i(t)}.$$

By the arguments given in the proofs of Theorems 3.1 and 5.1, we can show that $n^{1/2}\{\hat{\Lambda}_{\text{opt}}(\hat{\theta}_{\text{opt}}, t_0) - \Lambda_0(t_0)\} \to_L \mathscr{N}(0, v_0)$. The details can be found in Lin and Ying (1993).

REFERENCES

AALEN, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statist.* **2** 1–25. Springer, New York.

AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8** 907–925.

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

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

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

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

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

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.

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

GOFFMAN, C. (1965). *Calculus of Several Variables.* Harper and Row, New York.

HUFFER, F. W. and MCKEAGUE, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Amer. Statist. Assoc.* **86** 114–129.

LIN, D. Y., FLEMING, T. R. and WEI, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81** 73–81.

LIN, D. Y. and YING, Z. (1993). Semiparametric analysis of general additive–multiplicative intensity models for counting processes. Technical Report 129, Dept. Biostatistics, Univ. Washington.

LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71.

MCKEAGUE, I. W. and SASIENI, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81** 501–514.

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

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

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

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

YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184.

DEPARTMENT OF BIOSTATISTICS, SC-32
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENT OF STATISTICS
HILL CENTER, BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903