The Annals of Statistics
1990, Vol. 18, No. 1, 329–353

NONPARAMETRIC SURVIVAL ANALYSIS WITH TIME-DEPENDENT COVARIATE EFFECTS: A PENALIZED PARTIAL LIKELIHOOD APPROACH

BY DAVID M. ZUCKER[1] AND ALAN F. KARR[2]

The Johns Hopkins University

Techniques are developed for nonparametric analysis of data under a Cox-regression-like model permitting time-dependent covariate effects determined by a regression function $\beta_0(t)$. Estimators resulting from maximization of an appropriate penalized partial likelihood are shown to exist and a computational approach is outlined. Weak uniform consistency (with a rate of convergence) and pointwise asymptotic normality of the estimators are established under regularity conditions. A consistent estimator of a common baseline hazard function is presented and used to construct a consistent estimator of the asymptotic variance of the estimator of the regression function. Extensions to multiple covariates, general relative risk functions and time-dependent covariates are discussed.

1. The model. In this paper we consider regression analysis of censored survival data in a setting allowing for time-varying covariate influences, within the context of a particular model for the covariate-specific hazard function $\lambda(t|\mathbf{z})$. More precisely, we stipulate that conditional on the p-vector covariate value $\mathbf{z}$,

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\exp\left[\sum_{j=1}^{p} \beta_{0j}(t) z_j\right], \tag{1.1}$$

where $\boldsymbol{\beta}_0(t)$ is an unknown function taking values in $\mathbf{R}^p$ and λ_0 is an unknown, nonnegative function. Only smoothness assumptions are imposed on $\boldsymbol{\beta}_0$ and λ_0. Note that the covariates, albeit random, are independent of time (but see Section 10), while their effects on the hazard rate do depend on time.

Evidently (1.1) generalizes the now-famous Cox regression model [Cox (1972)]; however, a key feature of that model—proportionality of hazard functions for individuals with different covariate values—is lost. Several authors have considered variable-influence covariates in models similar to (1.1), but only under rather stringent assumptions on the functional form of $\boldsymbol{\beta}_0$. Cox (1972) and others consider, for example, the case of polynomial $\boldsymbol{\beta}_0$; Brown (1975), in an analogous discrete-time model, takes $\boldsymbol{\beta}_0$ to be a step function. These and related models are in fact equivalent, via redefinition of the covariates, to a version of the ordinary

Received October 1986; revised March 1989.

[1] Current address: National Heart, Lung and Blood Institute, Bethesda, Md. 20892.

[2] Research supported in part by Air Force Office of Scientific Research, AFSC, under grant 82-0029. The United States Government is authorized to reproduce and distribute reprints for government purposes.

AMS 1980 *subject classifications.* 60G05, 62G05, 62J02, 62M09.

Key words and phrases. Survival analysis, covariate, Cox regression model, penalized maximum likelihood estimation, partial likelihood, consistency, asymptotic normality.

Cox model with time-dependent covariates that may be random but are known fully at time zero. For our model, no such equivalence obtains.

Aalen (1980) studied a model for nonparametric regression analysis of counting processes that, when specialized to our setting (covariates fixed in time; at most one failure per individual), yields the model

$$\lambda(t|\mathbf{z}) = \lambda_0(t) + \sum_{j=1}^{p} \beta_{0j}(t) z_j. \tag{1.2}$$

Aalen considered estimation of integrals of β_{0j}, which are difficult to interpret physically, and only sketched a derivation of asymptotic properties. McKeague (1986) investigated (1.2) further, within the context of linear regression analysis for semimartingales, and discussed properties of least-squares estimators of $\boldsymbol{\beta}_0$ constructed via the method of sieves [Grenander (1981)]. Both authors rely heavily on linear structure present in (1.2) but absent from (1.1).

Before proceeding, we introduce notation, assumptions and stochastic processes of interest. Associated with each individual i is a triplet $(T_i^0, V_i, \mathbf{Z}_i)$, in which T_i^0 is a nonnegative random variable representing that individual's (potential) failure time, V_i is the (potential) censoring time and $\mathbf{Z}_i$ is a random p-vector of covariates. The observable data for individual i are $T_i = \min\{T_i^0, V_i\}$, $D_i = 1(T_i = T_i^0)$ and $\mathbf{Z}_i$. We assume throughout that

1. The $(T_i^0, V_i, \mathbf{Z}_i)$ are i.i.d. copies of a triplet $(T^0, V, \mathbf{Z})$, in which T^0 and V are conditionally independent given the covariate $\mathbf{Z}$.
2. The distributions of T^0 and V are absolutely continuous.
3. The covariate $\mathbf{Z}$ is bounded; without loss of generality we rescale $\mathbf{Z}$ so that each component lies in the interval $[0, 1]$.

We denote by $\lambda(t|\mathbf{z})$ the covariate-conditional hazard function of T^0,

$$\lambda(t|\mathbf{z}) = \lim_{h \downarrow 0} \frac{1}{h} P\{T^0 \le t + h | T^0 > t, \mathbf{Z} = \mathbf{z}\},$$

whose existence is presumed. Inference for $\lambda(t|\mathbf{z})$ under the model (1.1) is the principal subject of the paper.

We further define $N_i(t) = 1(T_i \le t, D_i = 1)$, the counting process of observed failures—there is at most one—for individual i; $Y_i(t) = 1(T_i \ge t)$, the "at risk" indicator process; $\lambda_i(t) = \lambda(t|\mathbf{Z}_i)Y_i(t)$, the stochastic intensity for the counting process N_i; $M_i(t) = N_i(t) - \int_0^t \lambda_i(u)\,du$, the innovation martingale. Also, let $\overline{N}(t) = (1/n)\sum_1^n N_i(t)$, with $\overline{M}(t)$ defined analogously.

The remainder of this paper is organized in the following manner. Our emphasis is on estimation of $\boldsymbol{\beta}_0$. Section 2 contains introductory discussion concerning our estimators of $\boldsymbol{\beta}_0$, which are obtained by maximization of a penalized analogue of the partial likelihood function central to analysis of the ordinary Cox model. Section 3 collects requisite notation, assumptions and preliminaries. Existence, characterization and computation of estimators are addressed in Section 4. In Section 5 we present a series of lemmas needed for our treatment of asymptotics. Sections 6 and 7 treat consistency, with respect to the uniform norm and asymptotic normality, respectively, of our estimators. In Section 8, we consider estimation of the baseline hazard function λ_0. Section 9 discusses estimation of the asymptotic variance of the estimator for β_0. In these sections, to simplify the presentation and notation, we assume that the covariate is one-dimensional. Section 10 contains remarks on the multidimensional case and extensions to the Prentice and Self (1983) generalization of the Cox model and to time-dependent covariates.

The reader should note that, by comparison with the ordinary Cox model, which provides a concise description of survival data, the model (1.1) attempts to provide more detailed information on covariate effects, but at the cost of added complexity. As a consequence, whereas the Cox model can be applied effectively to data sets of moderate size, the techniques described in this paper apply only to large sample sizes.

2. Estimating the regression parameter function. The main problem of interest is estimating the regression parameter function β_0, which we propose to do by maximizing a "penalized" version of the partial likelihood used as the basis for statistical inference for the ordinary Cox regression model. More precisely, the proposed estimator—referred to as the maximum penalized partial likelihood estimator (MPPLE)—is the maximizer of

$$(2.1)\quad L(\beta) = \frac{1}{n}\sum_{i=1}^{n} D_i\left[\beta(T_i)Z_i - \log\left(\sum_{j=1}^{n} Y_j(T_i)e^{\beta(T_i)Z_j}\right)\right] - \frac{1}{2}\alpha_n[\beta,\beta].$$

Here, α_n are positive numbers chosen by the statistician and, with $m \geq 3$ an integer also chosen by the statistician,

$$(2.2)\qquad [f,g] = \int_0^1 f^{(m)}(t)g^{(m)}(t)\,dt$$

for f and g belonging to the Sobolev space $H^m = H^m[0,1]$ of piecewise m-times differentiable functions f with $[f,f] < \infty$. The first term in (2.1) is (except for the factor $1/n$, added for convenience) exactly the logarithm of the Cox partial likelihood. The second term in (2.1) is a penalty funtional designed to make the estimator smooth and thereby reduce variance. Also, if $n > 1$ and for any observed ($D_i = 1$) failure, either $Z_i = \min\{Z_j\colon Y_j(T_i^0) = 1\}$ or $Z_i = \max\{Z_j\colon Y_j(T_i^0) = 1\}$, then (2.1) without the penalty has no maximizer: The unpenalized log-likelihood can be made arbitrarily large by making $|\beta(T_i^0)|$ large. For the estimators to be consistent (Section 6), it is necessary that $\alpha_n \to 0$, but at a controlled rate, as $n \to \infty$.

The idea of maximizing a penalized likelihood to obtain a nonparametric estimator goes back, in the context of estimating the probability density function f associated with i.i.d. random variables $X_1, \dots, X_n$, to Good and Gaskins (1971). In this situation, the penalized log-likelihood is given by $\sum_{i=1}^n \log f(X_i) - \alpha_n\Phi(f)$, where $\Phi(f)$ is a "flamboyance functional" and (α_n) is a sequence of positive numbers converging to zero as $n \to \infty$. Maximization is typically restricted to nonnegative functions integrating to unity; for our problem, there are no corresponding constraints.

Silverman (1982) considered the case $\Phi(f) = [\log f, \log f]$, with $[f, g]$ defined by (2.2), and for this estimator established existence, consistency in probability and asymptotic normality. Silverman's key idea is to introduce a modified version of the penalized log-likelihood depending on the unknown f; to show, using orthonormal expansions and Hilbert space theory, that the maximizer of the modified likelihood converges to the true log density; and finally to show that the maximizer of the modified likelihood also converges to the maximizer of the original likelihood. We rely heavily on this approach in establishing properties of our estimators.

3. Notation, assumptions and preliminaries.

Notation. We use the following notation, where $x \in \mathbf{R}$:

$$S_p(x; s) = \frac{1}{n} \sum_{i=1}^{n} Y_i(s) Z_i^p e^{xZ_i}, \qquad p = 0, \dots, 3,$$

$$A(x; s) = \frac{S_1(x; s)}{S_0(x; s)},$$

$$V(x; s) = \frac{(1/n)\Sigma_{i=1}^n Y_i(s)[Z_i - A(x; s)]^2 e^{xZ_i}}{S_0(x; s)} = \frac{S_2(x; s)}{S_0(x; s)} - A(x; s)^2,$$

$$C(x; s) = \frac{(1/n)\Sigma_{i=1}^n Y_i(s)[Z_i - A(x; s)]^3 e^{xZ_i}}{S_0(x; s)},$$

$$s_p(x; s) = E[Y(s) Z^p e^{xZ}], \qquad p = 0, \dots, 3,$$

$$a(x; s) = \frac{s_1(x; s)}{s_0(x; s)},$$

$$v(x; s) = \frac{E[Y(s)(Z - a(x; s))^2 e^{xZ}]}{s_0(x; s)} = \frac{s_2(x; s)}{s_0(x; s)} - a(x; s)^2,$$

$$c(x; s) = \frac{E[Y(s)(Z - a(x; s))^3 e^{xZ}]}{s_0(x; s)},$$

$$v_0 = \frac{1}{2} \inf_s v(\beta_0(s); s),$$

$$V_1(x; s) = \max\{V(x; s), v_0\},$$

$$v_1(x; s) = \max\{v(x; s), v_0\},$$

$$w(s) = \lambda_0(s) s_0(\beta_0(s); s) v(\beta_0(s); s).$$

Here and throughout β_0 denotes the "true" value of the unknown parameter; expectations above are taken under this value.

Recalling the assumptions that $0 \le Z \le 1$, it is clear that $0 \le A(x; s) \le 1$ and $0 \le V(x; s) \le 1$. Also, by direct computation, $(d/dx)A(x; s) = V(x; s)$, $(d/dx)V(x; s) = C(x; s)$ and $|C(x; s)| \le 1$. Similarly, $0 \le s_p(x; s) \le e^{|x|}$ for each p, $0 \le v(x; s) \le 1$, and $w(s) \ge 0$. Moreover, $(d/dx)s_p(x; s) = s_{p+1}(x; s)$ for $p = 0, 1, 2$, $(d/dx)a(x; s) = v(x; s)$ and $(d/dx)v(x; s) = c(x; s)$; differentiation within expectations is justified by the dominated convergence theorem.

We write $S_p(\beta, s)$ for $S_p(\beta(s); s)$, with similar abbreviations of the other quantities defined above.

Assumptions. The following assumptions are in force for the remainder of the paper.

ASSUMPTION A. The parameters α_n are deterministic.

ASSUMPTION B. $w(s) \ge w_0$ for some positive constant w_0.

ASSUMPTION C. $\beta_0 \in H^m$.

Note that Assumption A prevents data-dependent choice of the α_n. Assumption B is mild, and ensures that there is adequate "action" on the entire interval $[0, 1]$, in terms of failures and covariate variability, for estimation of β_0 to be meaningful.

Our proof of asymptotic normality requires an additional technical assumption.

ASSUMPTION D. w is $(2m - 1)$ times continuously differentiable on $[0, 1]$.

This assumption is satisfied if β_0, λ_0 and $s \to P\{V \ge s | Z = z\}$ are $(2m - 1)$ times continuously differentiable.

Note that m is chosen by the statistician, who can tailor the choice to whatever smoothness assumptions seem appropriate. Some smoothness, however, is necessary in order that our results hold: We require $m \ge 3$ for consistency and $m \ge 4$ for asymptotic normality.

Finally, up to quantities not depending on β, the negative of the log penalized likelihood is given by

$$(3.1)\quad H(\beta) = \frac{1}{2}\alpha[\beta, \beta] + \frac{1}{n}\sum_{i=1}^{n} D_i\left(\log\frac{S_0(\beta, T_i)}{S_0(\beta_0, T_i)} - Z_i[\beta(T_i) - \beta_0(T_i)]\right).$$

(*Note*: From this point on, dependence of α on n will generally be suppressed from the notation.) Direct computation shows that the first- and second-order

Gâteaux differentials of $H(\beta)$ are given for $f, g \in H^m$ by

$$\delta H(\beta; f) = \alpha[\beta, f] + \frac{1}{n}\sum_{i=1}^{n} D_i[A(\beta, T_i) - Z_i]\, f(T_i), \tag{3.2}$$

$$\delta^2 H(\beta; f, g) = \alpha[f, g] + \frac{1}{n}\sum_{i=1}^{n} D_i V(\beta, T_i) f(T_i) g(T_i). \tag{3.3}$$

4. Existence and characterization of MPPLE. Define, for $f, g \in H^m$,

$$\langle f, g\rangle_* = \sum_{j=1}^{m-1} f^{(j)}(0) g^{(j)}(0) + \int_0^1 f^{(m)}(t) g^{(m)}(t)\, dt.$$

Then $\langle f, g\rangle_*$ is an inner product on H^m whose induced norm is equivalent to the norm induced by the standard inner product on H^m. Because point evaluation is a continuous linear functional on H^m, for each $t \in [0,1]$ there exists a Riesz representer function $k_t \in H^m$ such that $\langle k_t, f\rangle_* = f(t)$ for every $f \in H^m$.

With this background, the main result of the section is as follows.

THEOREM 1. (a) *If $L(\beta)$ has a maximizer, then there exists a maximizer that lies in the finite-dimensional subspace of H^m spanned by the $(m-1)$-degree polynomials and the functions $\{k_{T_i}: D_i = 1\}$.*

(b) *There exists a maximizer of $L(\beta)$ if* (i) *there exists a solution to the problem of maximizing $L(\beta)$ with β restricted to be an $(m-1)$-degree polynomial and* (ii) *for every pair β, h of $(m-1)$-degree polynomials,*

$$\int_0^1 V(\beta, s) h^2(s)\, d\bar{N}(s) > 0. \tag{4.1}$$

(c) *With probability* 1, *the existence conditions in* (b) *are fulfilled for all sufficiently large n.*

PROOF. Part (a) follows using the argument employed in O'Sullivan, Yandell and Raynor (1986) to prove a corresponding result for a penalized maximum likelihood estimator in the context of a generalized linear model [Nelder and Wedderburn (1972)]; (b) follows by reasoning similar to that in Silverman [(1982), Theorem 4.1]. It remains to prove (c).

To verify that (b)(i) holds for n large, consider the $(m-1)$-degree polynomial $\beta(s) = \sum_{l=0}^{m-1} b_l s^l$ (which satisfies $[\beta, \beta] = 0$) and write

$$L(\beta) = \left(\prod_{i=1}^{n}\left[\sum_{j=1}^{n} Y_j(T_i)\exp\left\{(Z_j - Z_i)\sum_{l=0}^{m-1} b_l T_i^l\right\}\right]^{D_i}\right)^{-1}.$$

To show that a maximizing $(m-1)$-degree polynomial exists, it suffices to show that for every unit vector δ, the function $\beta_\gamma(s) = \gamma\Sigma\delta_l s^l$ satisfies $L(\beta_\gamma) \to 0$ as $|\gamma| \to \infty$. This, in turn, holds if for each i with $D_i = 1$, $(Z_{j_1} - Z_i)\Sigma\delta_l T_i^l \le 0$ for some j_1 and $(Z_{j_2} - Z_i)\Sigma\delta_l T_i^l \ge 0$ for some j_2 such that $T_{j_1} < T_i$ and $T_{j_2} < T_i$ and in addition there exist i_1^* for which the j_1-inequality is strict and i_2^* for which the j_2-inequality is strict. With probability 1, these conditions hold for n sufficiently large.

To verify (b)(ii), note that for n large, there exist individuals $i_1, \dots, i_m$ with distinct and actually observed failure times $T_{i_1}, \dots, T_{i_m}$ and individuals $j_1, \dots, j_m$ (failed or censored) such that $T_{j_r} > T_{i_r}$ and $Z_{j_r} \neq Z_{i_r}$ for each r. Evidently $V(\beta, T_{i_r}) > 0$ for every β; (b)(ii) then follows because an $(m-1)$-degree polynomial can have at most $m-1$ distinct roots. □

By standard variational arguments, for each t,

$$k_t(s) = \sum_{j=0}^{2m-1} c_j (s-t)^j + \frac{(s-t)_+^{2m-1}}{(2m-1)!},$$

where $x_+ = \max\{x, 0\}$ and the c_j are known constants. By this representation and Theorem 1(a), the problem of computing the MPPLE reduces to a finite-dimensional maximization.

5. Preliminary lemmas. This section contains preliminary lemmas needed to develop asymptotic properties of the MPPLE $\hat{\beta}$. The first results pertain to asymptotic behavior of the quantities $S_p(\beta_0, s)$, $p = 0, 1, 2$, defined in Section 3. Let $D[0,1]$ denote the set of right-continuous, left-limited functions on $[0,1]$; weak convergence of stochastic processes taking values in $D[0,1]$ is defined in terms of the Skorohod topology [Billingsley (1968)].

LEMMA 1. *For $p = 0, 1, 2$ there exists a continuous, mean-zero Gaussian process $G_p(t)$, with variance function that of the process $(Y(1-t)Z^p e^{\beta_0(1-t)Z})$, such that as $n \to \infty$,*

$$\left(\sqrt{n}\left[S_p(\beta_0, 1-t) - s_p(\beta_0, 1-t)\right]\right)_{0 \le t \le 1} \to_d G_p.$$

PROOF. Let $W_p(t) = Y(1-t)Z^p e^{\beta_0(1-t)Z} - s_p(\beta_0, 1-t)$. Then $W_p \in D[0,1]$ and $E[W_p(t)] < \infty$ for all t. We employ the central limit theorem of Hahn [(1978), Theorem 2], for which we must show that:

(i) There are constants $\xi_{1p} > \frac{1}{2}$ and nondecreasing, continuous functions B_{1p} such that $E[(W_p(u) - W_p(t))^2] \le [B_{1p}(u) - B_{1p}(t)]^{\xi_{1p}}$ for $u \ge t$.

(ii) There exist constants $\xi_{2p} > 1$ and nondecreasing, continuous functions B_{2p} such that for $s \le t \le u$,

$$E\left[\left(W_p(u) - W_p(t)\right)^2 \left(W_p(t) - W_p(s)\right)^2\right] \le \left[B_{2p}(u) - B_{2p}(s)\right]^{\xi_{2p}}.$$

Straightforward manipulations show that these conditions are satisfied by $\xi_{1p} = 1$, $\xi_{2p} = 2$ and $B_{lp}(u) = K_{lp}[u + P\{1 - u \le T < 1\}]$ for suitable constants K_{lp}. □

LEMMA 2. *As* $n \to \infty$,

$$\sup_s |S_p(\beta_0, s) - s_p(\beta_0, s)| = O_P(n^{-1/2}), \qquad p = 0,1,2, \tag{5.1}$$

$$\sup_s |A(\beta_0, s) - a(\beta_0, s)| = O_P(n^{-1/2}), \tag{5.2}$$

$$\sup_s |V(\beta_0, s) - v(\beta_0, s)| = O_P(n^{-1/2}). \tag{5.3}$$

PROOF. These follow directly from Lemma 1 and the continuous mapping theorem [Billingsley (1968), Theorem 5.2]. □

The final preliminary lemma pertains to asymptotic behavior of $\overline{N}(t)$.

LEMMA 3. *As* $n \to \infty$,

$$\sup_t \left| \overline{N}(t) - \int_0^t \lambda_0(s) s_0(\beta_0, s)\, ds \right| = O_P(n^{-1/2}). \tag{5.4}$$

PROOF. Defining $\overline{M}(t) = \overline{N}(t) - \int_0^t \lambda_0(s) S_0(\beta_0, s)\, ds$, we have

$$\sup_t \left| \overline{N}(t) - \int_0^t \lambda_0(s) s_0(\beta_0, s)\, ds \right| \le \left(\int_0^1 \lambda_0(s)\, ds \right) \sup_t |S_0(\beta_0, t) - s_0(\beta_0, t)| + \sup_t |\overline{M}(t)|.$$

The first term is $O_P(n^{-1/2})$ by Lemma 2; the second may be shown to be $O_P(n^{-1/2})$ using the martingale central limit theorem of Rebolledo (1980), as in Andersen and Gill (1982). □

6. Consistency of the MPPLE. In this section, the MPPLE $\hat{\beta}$ is shown to converge to β_0 in probability with respect to the uniform norm on $[0,1]$; the argument is patterned after Silverman (1982). Define

$$H_1(\beta) = \tfrac{1}{2}\alpha[\beta, \beta] + \tfrac{1}{2}\int_0^1 w(s)[\beta(s) - \beta_0(s)]^2\, ds - (1/n) \sum_{i=1}^n \int_0^1 [Z_i - A(\beta_0, s)][\beta(s) - \beta_0(s)]\, dN_i(s). \tag{6.1}$$

Then the idea of the proof is this:

1. Show that the minimizer β_1 of $H_1(\cdot)$ converges to β_0 as $n \to \infty$.
2. Show that for n sufficiently large, β_1 is close to $\hat{\beta}$.

The motivation for H_1 is as follows. Starting with (3.1) for $H(\beta)$, a two-term Taylor series expansion suggests the "approximation"

$$H(\beta) \doteq \tfrac{1}{2}\alpha[\beta, \beta] + \tfrac{1}{2}\int_0^1 V(\beta_0, s)[\beta(s) - \beta_0(s)]^2\, d\overline{N}(s) - (1/n) \sum_{i=1}^n \int_0^1 [Z_i - A(\beta_0, s)][\beta(s) - \beta_0(s)]\, dN_i(s).$$

The results in Section 5 further suggest replacing $V(\beta_0, s)$ by $v(\beta_0, s)$ and $d\bar{N}(s)$ by $\lambda_0(s)s_0(\beta_0, s)\,ds$. This gives $H(\beta) \doteq H_1(\beta)$.

It may be possible to establish a faster rate of convergence for $\hat{\beta}$ using developments in the smoothing spline literature; see Wahba (1985) and references therein. Also, it may be possible to use these developments to construct procedures for selecting the penalty weight α based on the observed data. These possibilities are not explored here.

To begin the consistency argument, define, for $f, g \in H^m$,

$$\langle f, g \rangle_w = \int_0^1 f(s)g(s)w(s)\,ds,$$

$$\langle f, g \rangle_{H^m} = \langle f, g \rangle_w + [f, g].$$

Then, by virtue of the assumption that $0 < w_0 \le w(s) \le \|w\|_\infty$ and Sobolev space theory [as in Silverman (1982)], there exist functions $(\phi_\nu)_{\nu=0}^\infty$ in H^m and numbers $1 = \mu_0 \ge \mu_1 \ge \mu_2 \ge \cdots \ge 0$ such that the sequence (ϕ_ν) is an orthonormal basis for $L^2[0,1]$ under the inner product $\langle f, g \rangle_w$ and $(\mu_\nu^{1/2}\phi_\nu)$ is an orthonormal basis for H^m under the inner product $\langle f, g \rangle_{H^m}$. In particular, with $\rho_\nu = \mu_\nu^{-1} - 1$,

$$\begin{aligned} \langle \phi_\nu, \phi_\eta \rangle_w &= \delta_{\nu\eta}, \\ \langle \phi_\nu, \phi_\eta \rangle_{H^m} &= \mu_\nu^{-1}\delta_{\nu\eta}, \\ [\phi_\nu, \phi_\eta] &= \rho_\nu \delta_{\nu\eta}. \end{aligned} \tag{6.2}$$

Let (b_ν) and $(b_{0\nu})$ be the coefficients in the expansions of β and β_0, respectively, in terms of (ϕ_ν). Then, using (6.2),

$$H_1(\beta) = \tfrac{1}{2}\alpha \sum_{\nu=0}^{\infty} \rho_\nu b_\nu^2 + \tfrac{1}{2} \sum_{\nu=0}^{\infty} (b_\nu - b_{0\nu})^2 - \sum_{\nu=0}^{\infty} X_\nu (b_\nu - b_{0\nu}),$$

where

$$X_\nu = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 [Z_i - A(\beta_0, s)]\,\phi_\nu(s)\,dN_i(s). \tag{6.3}$$

Evidently H_1 may be minimized by minimizing each term in the ν-summation individually. The coefficients of the minimizer β_1 are given by

$$b_{1\nu} = \frac{X_\nu + b_{0\nu}}{1 + \alpha\rho_\nu}. \tag{6.4}$$

The next lemma presents properties of the X_ν.

LEMMA 4. *An alternative expression for X_ν is*

$$X_\nu = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 [Z_i - A(\beta_0, s)]\,\phi_\nu(s)\,dM_i(s). \tag{6.5}$$

Consequently, $E[X_\nu] = 0$ *and* $\mathrm{Var}(X_\nu) \le 1/n$.

PROOF. Recall that $dN_i(s) = dM_i(s) + \lambda_0(s)Y_i(s)e^{\beta_0(s)Z_i}\,ds$; hence

$$X_\nu = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 [Z_i - A(\beta_0, s)]\phi_\nu(s)\,dM_i(s)$$
$$+ \int_0^1\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(s)Z_i e^{\beta_0(s)Z_i}\right)\phi_\nu(s)\lambda_0(s)\,ds$$
$$- \int_0^1 A(\beta_0, s)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(s)e^{\beta_0(s)Z_i}\right)\phi_\nu(s)\lambda_0(s)\,ds.$$

The term in parentheses in the second integral is $S_1(\beta_0, s)$; that in parentheses in the third integral is $S_0(\beta_0, s)$. Because $A(\beta_0, s) = S_1(\beta_0, s)/S_0(\beta_0, s)$, the second and third terms cancel, yielding (6.5). The remaining assertions follow from (6.5) using martingale theory [Andersen and Gill (1982)]. □

The proposition below gives probability bounds for the distance between β_1 and β_0 in the uniform and H^1 norms.

PROPOSITION 1. *There exist constants* $C_\varepsilon^{(1)}$, $\varepsilon > 0$, *and* $C^{(2)}$, *not depending on* n, *such that*

$$E\left[\|\beta_1 - \beta_0\|_\infty^2\right] \le C_\varepsilon^{(1)}\left[\alpha^{-\varepsilon/2m}\left(n^{-1}\alpha^{-1/m} + \alpha^{1-1/m}\right)\right], \tag{6.6}$$

$$E\left[\|\beta_1 - \beta_0\|_{H^1}^2\right] \le C^{(2)}\left[\alpha^{-1/2m}\left(n^{-1}\alpha^{-1/m} + \alpha^{1-1/m}\right)\right]. \tag{6.7}$$

In each of these bounds, the first term represents variance and the second term represents squared bias. They are proved using (6.4) and Lemma 4 in the manner of Silverman (1982).

Before turning to the difference between β_1 and $\hat{\beta}$, some further preliminaries are needed. Define, for arbitrary β,

$$\begin{aligned} H_M(\beta) &= \tfrac{1}{2}\alpha[\beta, \beta] + \int_0^1\int_{\beta_0(s)}^{\beta(s)}\int_{\beta_0(s)}^{u} V_1(x; s)\,dx\,du\,d\overline{N}(s) \\ &\quad - (1/n)\sum_{i=1}^{n}\int_0^1 [Z_i - A(\beta_0, s)][\beta(s) - \beta_0(s)]\,dN_i(s). \end{aligned} \tag{6.8}$$

By direct calculation, the first- and second-order Gâteaux differentials of H_M are given, for $f, g \in H^m$, by

$$\begin{aligned} \delta H_M(\beta; f) &= \alpha[\beta, f] + \int_0^1 f(s)\int_{\beta_0(s)}^{\beta_1(s)} V_1(x, s)\,dx\,d\overline{N}(s) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\int_0^1 f(s)[Z_i - A(\beta_0, s)][\beta(s) - \beta_0(s)]\,dN_i(s) \end{aligned} \tag{6.9}$$

and

$$\delta^2 H_m(\beta; f, g) = \alpha[f, g] + \int_0^1 f(s)g(s)V_1(\beta, s)\, d\bar{N}(s)$$
$$\geq \alpha[f, g] + v_0 \int_0^1 f(s)g(s)\, d\bar{N}(s). \tag{6.10}$$

Because β_1 minimizes H_1, for each $f \in H^m$,

$$0 = \delta H_1(\beta_1; f)$$
$$= \alpha[\beta_1, f] + \int_0^1 w(s)[\beta_1(s) - \beta_0(s)]\, f(s)\, ds$$
$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^1 [Z_i - A(\beta_0, s)]\, dN_i(s).$$

In conjunction with (6.9), this implies that

$$\delta H_M(\beta_1; f) = \int_0^1 f(s) \int_{\beta_0(s)}^{\beta_1(s)} V_1(x; s)\, dx\, d\bar{N}(s)$$
$$- \int_0^1 w(s)[\beta_1(s) - \beta_0(s)]\, f(s)\, ds. \tag{6.11}$$

The next proposition gives a probability bound for $\delta H_M(\beta_1; f)$.

PROPOSITION 2. *For $f \in H^m$ and possibly random,*

$$|\delta H_m(\beta_1; f)| = O_P\big(\big[\|\beta_1 - \beta_0\|_\infty^2 + n^{-1/2}\|\beta_1 - \beta_0\|_{H^1}\big]\|f\|_{H^1}\big), \tag{6.12}$$

where $\|g\|_{H^1}$ denotes the H^1-norm.

PROOF. By (6.11),

$$\delta H_M(\beta_1; f) = \int_0^1 f(s) \int_{\beta_0(s)}^{\beta_1(s)} [V_1(x; s) - v_1(x; s)]\, dx\, d\bar{N}(s)$$
$$+ \int_0^1 f(s) \int_{\beta_1(s)}^{\beta_0(s)} [v_1(x; s) - v(\beta_0, s)]\, dx\, d\bar{N}(s) \tag{6.13}$$
$$+ \int_0^1 f(s)[\beta_1(s) - \beta_0(s)]\, v(\beta_0, s)\big[d\bar{N}(s) - s_0(\beta_0, s)\lambda_0(s)\, ds\big].$$

These three terms will be treated in turn.

(i) For fixed s and for x between $\beta_0(s)$ and $\beta_1(s)$, the mean value theorem implies that

$$V(x; s) - v(x; s) = [V(\beta_0, s) - v(\beta_0, s)] + [C(x^*; s) - c(x^*; s)][x - \beta_0(s)]$$

for some x^* between $\beta_0(s)$ and $\beta_1(s)$. Now $|C(x^*; s) - c(x^*; s)| \leq 2$. Also, by Lemma 2, $\sup_s |V(\beta_0, s) - v(\beta_0, s)| = O_P(n^{-1/2})$ and therefore

$$|V_1(x; s) - v_1(x; s)| \leq |V(x; s) - v(x; s)| \leq O_P(n^{-1/2}) + 2\|\beta_1 - \beta_0\|_\infty.$$

Consequently,

$$\left|\int_0^1 f(s)\int_{\beta_0(s)}^{\beta_1(s)}[V_1(x;s) - v_1(x;s)]\,dx\,d\overline{N}(s)\right|$$

$$= O_P\big(\big[\|\beta_1 - \beta_0\|_\infty^2 + n^{-1/2}\|\beta_1 - \beta_0\|_\infty\big]\|f\|_{H^1}\big).$$

(ii) Fix $s \in [0,1]$ and x between $\beta_0(s)$ and $\beta_1(s)$. Then, by a mean value theorem argument as in (i), $|v(x;s) - v(\beta_0, s)| \le \|\beta_1 - \beta_0\|_\infty$. Now

$$v(x;s) \ge v(\beta_0, s) - \|\beta_1 - \beta_0\|_\infty \ge 2v_0 - \|\beta_1 - \beta_0\|_\infty$$

and so

$$|v_1(x;s) - v(x;s)| \le 2\cdot 1(\|\beta_1 - \beta_0\|_\infty \le v_0) \le 2v_0^{-1}\|\beta_1 - \beta_0\|_\infty.$$

Accordingly,

$$|v_1(x;s) - v(\beta_0, s)| \le \big(2v_0^{-1} + 1\big)\|\beta_1 - \beta_0\|_\infty,$$

so that

$$\left|\int_0^1 f(s)\int_{\beta_0(s)}^{\beta_1(s)}[v_1(x;s) - v(\beta_0,s)]\,dx\,d\overline{N}(s)\right| \le \big(2v_0^{-1} + 1\big)\|\beta_1 - \beta_0\|_\infty^2\|f\|_{H^1}.$$

(iii) The integrand in the third term in (6.13) is not predictable, so martingale theory cannot be applied; an alternative argument is needed. By integration by parts,

$$\left|\int_0^1 v(\beta_0, s)[\beta_1(s) - \beta_0(s)]\big[d\overline{N}(s) - s_0(\beta_0, s)\lambda_0(s)\,ds\big]\right|$$

$$(6.14)\qquad \begin{aligned} &\le v(\beta_0,1)|\beta_1(1) - \beta_0(1)|\,|f(1)|\left|\overline{N}(1) - \int_0^1 s_0(\beta_0,s)\lambda_0(s)\,ds\right| \\ &\quad + \sup_t\left|\overline{N}(t) - \int_0^t s_0(\beta_0,s)\lambda_0(s)\,ds\right| \\ &\qquad\times\int_0^1\left|\frac{d}{ds}[v(\beta_0,s)\lambda_0(s)[\beta_1(s) - \beta_0(s)]\,f(s)]\right|ds. \end{aligned}$$

By Lemma 3,

$$\sup_t\left|\overline{N}(t) - \int_0^t s_0(\beta_0,s)\lambda_0(s)\,ds\right| = O_P(n^{-1/2}).$$

Thus the first summand in (6.14) is $O_P(n^{-1/2}\|\beta_1 - \beta_0\|_{H^1}\|f\|_{H^1})$. The second summand in (6.14) is bounded by the product of a quantity that is $O_P(n^{-1/2})$

and

$$\int_0^1 \left| \frac{d}{ds} [v(\beta_0, s)[\beta_1(s) - \beta_0(s)] f(s)] \right| ds$$

$$= \int_0^1 |v'(\beta_0, s)[\beta_1(s) - \beta_0(s)] f(s)| \, ds$$

$$+ \int_0^1 |v(\beta_0, s)[\beta_1'(s) - \beta_0'(s)] f(s)| \, ds$$

$$+ \int_0^1 |v(\beta_0, s)[\beta_1(s) - \beta_0(s)] f'(s)| \, ds$$

$$\leq \max\{\|v'\|_\infty, 1\} \|\beta_1 - \beta_0\|_{H^1} \|f\|_{H^1}.$$

Combining (i), (ii) and (iii) completes the proof. □

REMARK. By an integration-by-parts argument similar to that used to prove (iii),

$$\left| \int_0^1 g^2(s) \, d\bar{N}(s) - \int_0^1 g^2(s) s_0(\beta_0, s) \lambda_0(s) \, ds \right| = O_P\left(n^{-1/2} \|g\|_{H^1}^2\right).$$

Consequently, for each β,

$$\delta^2 H_M(\beta; g, g) \geq 2v_0 \langle g, g \rangle + \alpha[g, g] - O_P\left(n^{-1/2} \|g\|_{H^1}^2\right).$$

But, with $g_\nu = \langle g, \phi_\nu \rangle$,

$$v_0 \langle g, g \rangle + \alpha[g, g] = \sum_{\nu=0}^{\infty} (2v_0 + \alpha\rho_\nu) g_\nu^2$$

$$= \sum_{\nu=0}^{\infty} \nu^{-2} (2v_0 + \alpha\rho_\nu) \nu^2 g_\nu^2$$

$$\geq \tilde{C} \alpha^{1/m} \sum_{\nu=0}^{\infty} \nu^2 g_\nu^2 \geq C^0 \alpha^{1/m} \|g\|_{H^1}^2,$$

for a suitable constant C^0, by Sobolev space theory. Here $\tilde{C} > 0$, as in Silverman [(1982), equation (7.5)]; to see the inequality involving $\tilde{C}$, note that

$$\alpha^{1/m} \nu^{-2} (1 + \alpha \nu^{2m}) = \left(1 + \alpha^{1/2} \nu^m\right)^2 / \left(\alpha^{1/2} \nu^m\right)^{2/m} \geq 1.$$

Therefore,

$$\delta^2 H_M(\beta; g, g) \geq C^0 \alpha^{1/m} \|g\|_{H^1}^2 - O_P\left(n^{-1/2} \|g\|_{H^1}^2\right). \tag{6.15}$$

The O_P in (6.15) is uniform in g. Thus, if $\alpha = O(n^{-\theta})$ with $\theta < m/2$, then the probability that H_M is uniformly convex converges to 1 as $n \to \infty$. When H_M is uniformly convex, it has a unique minimizer [Tapia and Thompson (1978), Appendix 1], which we denote by β_M.

By construction, $H_M(\beta) = H(\beta)$ if, for each s, $V(x; s) \geq v_0$ for every x between $\beta_0(s)$ and $\beta(s)$. Define the (compact) interval $I = [-1 - \inf_s \beta_0(s),$

$1 + \sup_s \beta_0(s)]$. Then by Theorem III.1 of Andersen and Gill (1982), applied with $Y_i(s)e^{xZ_i}$ viewed as a random element of the space of left-continuous, right-limited functions from $[0,1]$ to the set of continuous functions on I,

$$\sup\{|S_p(x;s) - s_p(x;s)|: s \in [0,1], x \in I\} \to 0 \quad \text{a.s. for } p = 0,1,2.$$

Consequently, for all n sufficiently large (depending on the realization ω), $V(x;s) \geq v_0$ for all $x \in I$ and $s \in [0,1]$.

Thus, provided $\|\beta - \beta_0\|_\infty \leq 1$, $H_M(\beta) = H(\beta)$, and similar equalities obtain for the first- and second-order Gâteaux differentials, for all n sufficiently large. Hence $\hat{\beta}$ will be equal to β_M if $\|\beta_M - \beta_0\|_\infty \leq 1$ and n is sufficiently large. Hence an analysis of the difference between β_M and β_1 provides information regarding that between $\hat{\beta}$ and β_1. This idea is implemented in the next proposition.

PROPOSITION 3. *Let $\varepsilon > 0$ be given and suppose that $\alpha_n = O(n^{-\theta})$ with $0 < \theta < 2m/(4+\varepsilon)$. Then provided $m \geq 3$,*

$$\|\hat{\beta} - \beta_1\|_{H^1} = O_P\left(n^{-1+[(4+\varepsilon)/2m]\theta} + n^{-[1-(4+\varepsilon)/2m]\theta} + n^{-[1+(1-7/2m)]\theta/2}\right). \tag{6.16}$$

REMARK. By the Sobolev embedding theorem, the same rate holds for $\|\hat{\beta} - \beta_1\|_\infty$.

PROOF. Put $g = \beta_1 - \beta_M$. Then, recalling that $\delta H_M(\beta_M; f) = 0$ for all $f \in H^m$, by Taylor's theorem for functionals [Graves (1927)],

$$\delta H_M(\beta_1; g) = \delta^2 H_M(\beta_M + \xi g; g, g) \tag{6.17}$$

for some $\xi \in [0,1]$. From (6.11),

$$\delta^2 H_M(\beta_M + \xi g; g, g) \geq C^0 \alpha^{1/m} \|g\|_{H^1}^2 - O_P\left(n^{-1/2}\|g\|_{H^1}^2\right). \tag{6.18}$$

On the other hand, Proposition 2 gives

$$|\delta H_M(\beta_1; g)| \leq O_P\left(\left[\|\beta_1 - \beta_0\|_\infty^2 + n^{-1/2}\|\beta_1 - \beta_0\|_{H^1}\right]\|g\|_{H^1}\right). \tag{6.19}$$

Putting (6.18) and (6.19) into (6.17) and cancelling a factor of $\|g\|_{H^1}$ yields, under the hypothesis regarding α_n,

$$\|\beta_M - \beta_1\|_{H^1} \leq \alpha^{-1/m} O_P\left(\|\beta_1 - \beta_0\|_\infty^2 + n^{-1/2}\|\beta_1 - \beta_0\|_{H^1}\right). \tag{6.20}$$

Therefore, by Proposition 1 and the hypothesis on α_n, and the Sobolev embedding theorem, $\|\beta_1 - \beta_0\|_\infty \to_P 0$, and for an appropriate constant c, $\|\beta_1 - \beta_M\|_\infty \leq c\|\beta_1 - \beta_M\|_{H^1} \to_P 0$, and so $\|\beta_M - \beta_0\|_\infty \to_P 0$. Accordingly, $P\{\beta_M = \hat{\beta}\} \to 1$.

Substituting the rates in Proposition 1 into (6.20), which now holds with β_M replaced by $\hat{\beta}$, completes the proof. □

Finally, Propositions 1 and 3 combine to yield the main result of this section.

THEOREM 2. *Let* $\varepsilon > 0$ *be given and suppose that* $\alpha_n = O(n^{-\theta})$ *with*

$$0 < \theta < \frac{2m}{4+\varepsilon}. \tag{6.21}$$

Then

$$\begin{aligned}\|\hat{\beta} - \beta_0\|_\infty = O_P\big(&n^{-1+[(4+\varepsilon)/2m]\theta} + n^{-1/2+[(2+\varepsilon)/4m]\theta}\\ &+n^{-[1+(1-7/2m)]\theta/2} + n^{-[1-(4+\varepsilon)/2m]\theta} + n^{-[1-(2+\varepsilon)/2m]\theta/2}\big).\end{aligned} \tag{6.22}$$

In this expression, the first two terms represent variability and the latter three represent bias.

7. Asymptotic normality of the MPPLE. In this section we establish pointwise asymptotic normality of the estimators $\hat{\beta}(t)$. Our treatment of asymptotic distribution theory differs from that of Silverman (1982): for his density estimators, he employed the theorem of Komlós, Major and Tusnády (1975) for strong approximation of the empirical distribution functions associated with a sequence of i.i.d. random variables by a Gaussian process. This argument does not carry over to our situation because there is no corresponding approximation for the i.i.d. sequence of martingales $\overline{M}_i(t) = \int_0^t [Z_i - A(\beta_0, s)]\, dM_i(s)$. Consequently, we were able to derive only a pointwise result.

We now introduce necessary notation:

$$\begin{aligned}
X_\nu^* &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 [Z_i - a(\beta_0, s)]\phi_\nu(s)\, dM_i(s),\\
\beta^*(t) &= \sum_{\nu=0}^{\infty} \frac{X_\nu^* + b_{0\nu}}{1+\alpha\rho_\nu}\phi_\nu(t),\\
\beta_\alpha(t) &= \sum_{\nu=0}^{\infty} \frac{b_{0\nu}}{1+\alpha\rho_\nu}\phi_\nu(t),\\
U(t) &= \sum_{\nu=0}^{\infty} \frac{X_\nu^*}{1+\alpha\rho_\nu}\phi_\nu(t),\\
R_\alpha(s,t) &= \sum_{\nu=0}^{\infty} \frac{1}{1+\alpha\rho_\nu}\phi_\nu(s)\phi_\nu(t),\\
r_\alpha(s,t) &= \sum_{\nu=0}^{\infty} \frac{1}{(1+\alpha\rho_\nu)^2}\phi_\nu(s)\phi_\nu(t).
\end{aligned} \tag{7.1}$$

With this notation and the expression (6.4) for the coefficients $b_{1\nu}$ of β_1, the MPPLE $\hat{\beta}$ satisfies

$$\begin{aligned}\hat{\beta}(t) - \beta_0(t) = &\left[\hat{\beta}(t) - \beta_1(t)\right] + \left[\beta_1(t) - \beta^*(t)\right]\\ &+\left[\beta_\alpha(t) - \beta_0(t)\right] + U(t).\end{aligned} \tag{7.2}$$

In this section, Assumption D is presumed in force. The idea of the proof is to show that $U(t)$, suitably standardized, is asymptotically normal and to establish appropriate bounds on the remaining terms on the right-hand side of (7.2). A bound on $\|\hat{\beta} - \beta_1\|_\infty$ was given in Proposition 3. The other steps of the proof of asymptotic normality are contained in the following series of lemmas.

LEMMA 5. *For each $\varepsilon > 0$ there exists a constant C_ε^* such that*

$$E\left[\|\beta_1 - \beta^*\|_\infty^2\right] \le C_\varepsilon^* n^{-2} a^{-(2+\varepsilon)/2m}. \tag{7.3}$$

PROOF. By martingale theory, $E[(X_\nu - X_\nu^*)^2] \le \kappa n^{-2}$ for some constant κ. Therefore, with $b_\nu^* = \langle \beta^*, \phi_\nu \rangle$, Sobolev space theory [Silverman (1982)] implies that

$$\begin{aligned} E\left[\|\beta_1 - \beta^*\|_\infty^2\right] &\le \sum_{\nu=0}^{\infty} \nu^{1+\varepsilon} E\left[(b_{1\nu} - b_\nu^*)^2\right] \\ &= C_\varepsilon \sum_{\nu=0}^{\infty} \frac{\nu^{1+\varepsilon} E\left[(X_\nu - X_\nu^*)^2\right]}{(1 + \alpha\rho_\nu)^2} \\ &\le C_\varepsilon \kappa n^{-2} \sum_{\nu=0}^{\infty} \frac{\nu^{1+\varepsilon}}{(1 + \alpha\rho_\nu)^2} \le C_\varepsilon^* n^{-2} a^{-(2+\varepsilon)/2m} \end{aligned}$$

for appropriate constants C_ε and C_ε^* independent of n. □

LEMMA 6. *For each $\varepsilon > 0$ there exists a constant $C_\varepsilon^* > 0$ such that*

$$\|\beta_\alpha - \beta_0\|_\infty^2 \le C_\varepsilon^* \alpha^{1-1/m-\varepsilon/2m}.$$

PROOF. This follows by a Sobolev space argument similar to that used to prove Lemma 5. □

PROPOSITION 4. *Provided that $n^{1/2}\alpha^{1/4m} \to \infty$, as $n \to \infty$,*

$$\frac{U(t)}{\sqrt{\operatorname{Var}(U(t))}} \to_d N(0,1). \tag{7.4}$$

PROOF. Put

$$W_{ni} = \int_0^1 R_\alpha(s,t)[Z_i - a(\beta_0, s)]\, dM_i(s). \tag{7.5}$$

Then evidently $U(t) = (1/n)\sum_{i=1}^n W_{ni}$ and the W_{ni} are i.i.d. as i varies with n fixed. Moreover, $E[W_{ni}] = 0$ by martingale theory and (making use of orthonormality of the ϕ_ν)

$$\sigma_n^2(t) = E\left[W_{ni}^2\right] = r_\alpha(t,t). \tag{7.6}$$

With t fixed, because the ϕ_ν are eigenfunctions of a differential operator [Naimark (1967)], there exist positive constants r_0 and r_1 (depending on t) such that

$$r_0\alpha^{-1/2m} \le \sigma_n^2 \le r_1\alpha^{-1/2m}. \tag{7.7}$$

By Chow and Teicher [(1978), Corollary 12.2.2], to prove (7.4) it suffices to show that for $\xi > 0$, $nP\{|W_{ni}|/\sigma_n > \xi n^{1/2}\} \to 0$ as $n \to \infty$. By Markov's inequality, for $\Delta > 0$,

$$nP\{|W_{ni}|/\sigma_n > \xi n^{1/2}\} \le \frac{E\left[|W_{ni}|^{2+\Delta}\right]}{\xi^{2+\Delta} n^{\Delta/2}\sigma_n^{2+\Delta}}. \tag{7.8}$$

The goal now is to show that the right-hand side of (7.8) converges to zero as $n \to \infty$.

As a preliminary, differential operator theory [Naimark (1967)] implies that $|\phi_\nu(t)|$ is bounded uniformly in ν and t, so that

$$\sup_{s,t} |R_\alpha(s, t)| \le C_\varepsilon^* \alpha^{-1/2m} \tag{7.9}$$

for some constant C_ε^*.

Now recall that $dM_i(s) = dN_i(s) - \lambda_0(s)Y_i(s)e^{\beta_0(s)Z_i}\,ds$. Hence, with $d|M_i(s)|$ denoting the total variation of the signed measure $dM_i(s)$,

$$d|M_i(s)| \le dN_i(s) + \|\lambda_0\|_\infty e^{\|\beta_0\|_\infty}\,ds.$$

Therefore

$$E\left[|W_{ni}|^{2+\Delta}\right] \le 2^{2+\Delta} E\left[\left(\int_0^1 |R_\alpha(s, t)|\,dN_i(s)\right)^{2+\Delta}\right] + \left(2\|\lambda_0\|_\infty e^{\|\beta_0\|_\infty}\right)^{2+\Delta}\left(\int_0^1 |R_\alpha(s, t)|\,ds\right)^{2+\Delta}$$

Because the counting process N_i has at most one point,

$$\begin{aligned} E\left[\left(\int_0^1 |R_\alpha(s, t)|\,dN_i(s)\right)^{2+\Delta}\right] &= E\left[\int_0^1 |R_\alpha(s, t)|^{2+\Delta}\,dN_i(s)\right] \\ &= \int_0^1 |R_\alpha(s, t)|^{2+\Delta}\lambda_0(s)s_0(\beta_0, s)\,ds \\ &\le \frac{\sup_{s,t}|R_\alpha(s, t)|^\Delta}{\inf_s v(\beta_0, s)} \int_0^1 R_\alpha^2(s, t)w(s)\,ds \\ &\le K_1\sigma_n^2\alpha^{-\Delta/2m}, \end{aligned}$$

where K_1 is a constant and where we have applied (7.3) and (7.9). By similar analysis, for a constant K_2,

$$\left(\int_0^1 |R_\alpha(s, t)|\,ds\right)^{2+\Delta} \le K_2\sigma_n^2\alpha^{-\Delta/2m}.$$

Thus for a suitable constant κ, $E[|W_{ni}|^{2+\Delta}] \le \kappa\sigma_n^2\alpha^{-\Delta/2m}$. By this last inequality, the right-hand side of (7.8) is bounded by $\xi^{-(2+\Delta)}\kappa(n^{1/2}\sigma_n\alpha^{1/2m})^{-\Delta}$ and this converges to zero as $n \to \infty$ by (7.7) and the hypothesis on α_n. □

At this point, the groundwork is laid for the main result of this section.

THEOREM 3. *Let $\varepsilon > 0$ be given and suppose that $\alpha_n = O(n^{-\theta})$ with*

$$\frac{1}{1-(1+\varepsilon)/2m} < \theta < \frac{2m}{7+\varepsilon}. \tag{7.10}$$

Also, suppose that $m \geq 4$. Then, as $n \to \infty$, for each fixed $t \in [0,1]$, with σ_n^2 defined in (7.6),

$$\sqrt{n/\sigma_n^2}\,[\hat{\beta}(t) - \beta_0(t)] \to_d N(0,1). \tag{7.11}$$

PROOF. By Slutsky's theorem, (7.2), and Proposition 4, it suffices to show that $(n/\sigma_n^2)^{1/2}[\hat{\beta}(t) - \beta_1(t)]$, $(n/\sigma_n^2)^{1/2}[\beta_1(t) - \beta^*(t)]$ and $(n/\sigma_n^2)^{1/2}[\beta_\alpha(t) - \beta_0(t)]$ converge in probability to zero as $n \to \infty$. Under (7.10), these follow from Proposition 1, Lemmas 5 and 6, and (7.7), respectively. □

Note that, by construction, choosing θ to satisfy (7.10) causes the squared bias of $\hat{\beta}(t)$ to converge to zero faster than the variance. This choice seems the most natural in terms of constructing confidence intervals for $\beta_0(t)$. On the other hand, to minimize the mean squared error of $\hat{\beta}(t)$, one should choose θ to balance the squared bias and variance. With this choice, Theorem 3 no longer applies; however, Proposition 4 remains valid for $U(t)$, which is the dominant term of those relating to variance.

8. Estimation of the baseline hazard function. The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ may be estimated by

$$\hat{\Lambda}(t) = \int_0^t \frac{1}{S_0(\hat{\beta}, s)}\, d\bar{N}(s).$$

This estimator, essentially a martingale estimator, is the analogue for the model (1.1) of the estimator commonly used to estimate the cumulative baseline hazard function in the ordinary Cox model. It is straightforward to show that under the conditions of Theorem 2, $\|\hat{\Lambda} - \Lambda_0\|_\infty$ converges in probability to zero with the rate given for $\|\hat{\beta} - \beta_0\|_\infty$ in (6.22).

To estimate λ_0 itself, one can estimate $\lambda^*(t) = \lambda_0(t)s_0(\beta_0, t)$ and then divide by $S_0(\hat{\beta}, t)$. That the functions S_0 have bounded derivatives and convergence of $\hat{\beta}$ to β_0 ensure that this approach will work provided that λ^* is estimated consistently. Several methods could be used to estimate λ^*. Probably the simplest is the kernel method, along the lines of Ramlau-Hansen (1983) and other authors. Alternatively, methods related to penalized likelihood could be employed; cf. Karr (1987) for sieve estimation in the multiplicative intensity model of Aalen (1980). The following theorem gives a formal result for the kernel method; it is proved by arguments analogous to those in Ramlau-Hansen (1983).

THEOREM 4. *Let Q be a function with bounded variation and support $[-1,1]$ such that $\int Q(t)\,dt = 1$ and $\int t^j Q(t)\,dt = 0$ for $j = 1, \dots, k-1$ for some* *integer $k \geq 1$. Let (γ_n) be a sequence of (nonrandom) positive constants with $\gamma_n = O(n^{-\rho})$. Suppose that λ^* is k times continuously differentiable. Define the estimator*

$$\hat{\lambda}^*(t) = \frac{1}{\gamma_n}\int_0^1 Q\left(\frac{t-s}{\gamma_n}\right)\frac{J(s)}{\overline{Y}(s)}\,d\overline{N}(s), \tag{8.1}$$

where $J(s) = 1(Y_i(s) = 1$ for some $i \leq n)$ and $\overline{Y}(s) = (1/n)\sum_{i=1}^n Y_i(s)$. Then for $0 < t_1 < t_2 < 1$,

$$\sup_{t\in[t_1,t_2]} |\hat{\lambda}^*(t) - \lambda^*(t)| = O_P(n^{-k\rho} + n^{-(1-\rho)/2}). \tag{8.2}$$

Because of the choice of Q, the estimator (8.1) may assume negative values; however, it may be modified slightly in order to be made positive without affecting the rate of convergence in (8.2).

In estimating the asymptotic variance $\sigma_n^2(t)$ of $\hat{\beta}(t)$ (Section 9), it is necessary to estimate $w(s) = \lambda^*(s)v(\beta_0, s)$, which may be done via

$$\hat{w}(s) = \hat{\lambda}^*(s)V(\hat{\beta}, s). \tag{8.3}$$

From Theorem 4, Theorem 2, and differentiability of V, the following result obtains.

PROPOSITION 5. *For $\hat{w}$ given by* (8.3) *and $0 < t_1 < t_2 < 1$,*

$$\sup_{t\in[t_1,t_2]} |\hat{w}(t) - w(t)| \to_P 0, \tag{8.4}$$

at the slower of the rate given in (6.22) *and that given in* (8.2).

In particular, if $k \geq \lceil m/2 \rceil$ and $\rho = 1/(2k+1)$, then (8.4) holds at the rate given for $\|\hat{\beta} - \beta_0\|_\infty$ in (6.22).

9. Estimation of the asymptotic variance. To apply Theorem 3, it is necessary to estimate the variance $\sigma_n^2(t)$ given by (7.6). That equation involves the quantity $R_\alpha(s, t)$, defined at the beginning of Section 7, which is also the reproducing kernel (R.K.) for the Hilbert space H^m under the inner product

$$\langle f, g\rangle_{H^m,\alpha} = \alpha[f, g] + \int_0^1 f(s)g(s)w(s)\,ds. \tag{9.1}$$

It is natural to estimate $\sigma_n^2(t)$ by substituting the estimator $\hat{w}$ of (8.3) for w in (9.1), computing the corresponding reproducing kernel $\hat{R}_\alpha$ and then substituting $\hat{R}_\alpha$ for R_α in (7.6). We show in this section that the resultant estimator $\hat{\sigma}_n^2(t)$ is, on a pointwise basis, consistent in the sense that as $n \to \infty$,

$$\frac{|\hat{\sigma}_n^2(t) - \sigma_n^2(t)|}{\sigma_n^2(t)} \to_P 0. \tag{9.2}$$

Additionally, an approach to computation of $\hat{\sigma}_n^2(t)$ is outlined.

The following result on reproducing kernels is the first step in demonstrating (9.2), as well as the basis of computational considerations.

PROPOSITION 6. *Suppose that* $\alpha > 0$ *is fixed and that* $f_1, f_2 \in H^m$ *are given. Define* R_i $(i = 1, 2)$ *to be the R.K. for* H^m *under the inner product*

$$\langle h_1, h_2 \rangle_{(i)} = \alpha[h_1, h_2] + \int_0^1 h_1(s) h_2(s) f_i(s)\, ds.$$

Then the following integral equation holds:

$$R_1(s,t) = R_2(s,t) + \int_0^1 [f_1(u) - f_2(u)] R_1(s,u) R_2(t,u)\, du. \tag{9.3}$$

PROOF. One simply calculates as follows:

$$\begin{aligned} R_1(s,t) &= \langle R_1(s,\cdot), R_2(t,\cdot)\rangle_{(2)} \\ &= \alpha[R_1(s,\cdot), R_2(t,\cdot)] + \int_0^1 f_2(u) R_1(s,u) R_2(t,u)\, du \\ &= \alpha[R_1(s,\cdot), R_2(t,\cdot)] + \int_0^1 f_1(u) R_1(s,u) R_2(t,u)\, du \\ &\quad + \int_0^1 [f_2(u) - f_1(u)] R_1(s,u) R_2(t,u)\, du \\ &= \langle R_1(s,\cdot), R_2(t,\cdot)\rangle_{(1)} + \int_0^1 [f_2(u) - f_1(u)] R_1(s,u) R_2(t,u)\, du \\ &= R_2(s,t) + \int_0^1 [f_2(u) - f_1(u)] R_1(s,u) R_2(t,u)\, du, \end{aligned}$$

which verifies (9.3). □

COROLLARY 1. *Let* $c_1 = \sup_{s,t}|R_1(s,t)|$ *and assume* $c_1\| f_2 - f_1\|_\infty < 1$. *Then*

$$\sup_{s,t}|R_1(s,t) - R_2(s,t)| \le \frac{c_1\|f_2 - f_1\|_\infty}{1 - c_1\|f_2 - f_1\|_\infty} c_1. \tag{9.4}$$

PROOF. By Proposition 6, $R_1(\cdot, t) = (I + \mathscr{B})R_2(\cdot, t)$, where $\mathscr{B}$ is the operator defined by

$$\mathscr{B}h(s) = \int_0^1 [f_2(u) - f_1(u)] R_1(s,u) h(u)\, du.$$

In view of the property that $\|\mathscr{B}\| \le c_1\|f_2 - f_1\|_\infty < 1$, operator theory implies that $(I + \mathscr{B})^{-1}$ exists and satisfies $\|(I + \mathscr{B})^{-1}\| \le 1/(1 - \|\mathscr{B}\|)$, from which (9.4) follows by routine computations. □

COROLLARY 2. *Suppose that the conditions of Proposition* 6 *and Corollary* 1 *hold and define*

$$r_i(s,t) = \int_0^1 R_1(s,u) R_2(u,t) f_i(u)\, du.$$

Then

$$\sup_{s,t}|r_1(s,t) - r_2(s,t)| \le c_1^2\|f_2 - f_1\|_\infty + (\|f_1\|_\infty + \|f_2 - f_1\|_\infty)$$

$$\times\left[\frac{2c_1^3\|f_2 - f_1\|_\infty}{1 - c_1\|f_2 - f_1\|_\infty} + \frac{c_1^4\|f_2 - f_1\|_\infty^2}{(1 - c_1\|f_2 - f_1\|_\infty)^2}\right]. \tag{9.5}$$

Consistency of $\hat{\sigma}_n^2(t)$ is shown by applying the development above with $f_1 = w$ and $f_2 = \hat{w}$, resulting in the following theorem.

THEOREM 5. *Under the conditions of Theorem* 2 *and the sentence following Proposition* 5, *and the additional stipulations that* $m \ge 4$ *and* $0 < \theta < 2m/(6 + \varepsilon)$, *where* $\varepsilon > 0$, *then* (9.2) *holds for each fixed t.*

Note that in this case, Theorem 4 holds with $\sigma_n^2(t)$ replaced by $\hat{\sigma}_n^2(t)$, so that one may thereby construct hypothesis tests and confidence intervals for $\beta_0(t)$ based on the normal distribution.

For computation of $\hat{\sigma}_n^2(t)$, the main idea is to use Proposition 6 with $f_1 = \alpha$ and $f_2 = \hat{w}$, and to solve (9.3) numerically. With $f_1 = \alpha$, $R_1 = \alpha^{-1}R^*$, where R^* is the R.K. for H^m under the *standard* inner product $[h_1, h_2] + \int h_1 h_2$, and must be determined (algebraically or numerically) only once. For fixed t, $R^*(s,t) = h_0(s)1(s \le t) + h_1(s)1(s > t)$, where h_0 and h_1 are linear combinations of the functions $\exp[\pi i l/m]$ (m odd) or $\exp[\pi i l/2m]$ (m even), with coefficients determined by a system of linear equations resulting from the conditions $h_0^{(j)}(0) = h_1^{(j)}(1) = 0$, $j = m+1, \dots, 2m$, $h_0^{(j)}(t) = h_1^{(j)}(t)$, $j = 0, \dots, m-1$ and $\int R^*(s,t)\,ds = 1$.

With R_1 determined, (9.3) may be solved for $R_2(s,t)$ using the Nystrom method [Delves and Mohamed (1985), Chapter 4], which involves approximating the integral by a sum calculated by standard numerical quadrature methods and based on partition points u_j. This leads to a system of linear equations for $R_2(u_j, t)$, which may be solved by matrix methods. Finally, the integral

$$\hat{\sigma}_n^2(t) = \int_0^1 R_2^2(u,t)\,du$$

may be approximated using the same partition and the values $R_2(u_j, t)$ substituted into the sum. When α is small, these procedures may have to be modified to account for machine precision.

10. Remarks on extensions. In this section we discuss briefly several extensions of our basic model.

The multivariate case. In the case that the covariate is a p-vector, the penalized log-likelihood is given by (2.1) with $\beta(T_i)Z_i$ replaced by $\boldsymbol{\beta}(T_i)^{\mathrm{T}}\mathbf{Z}_i$ (the superscript T denotes transpose) and $[f, g]$ defined for $f, g \in H^m[0,1]^p$ by

$$[f, g] = \sum_{j=1}^{p} a_j \int_0^1 f_j^{(m)}(t) g_j^{(m)}(t)\,dt,$$

where the a_j are fixed, positive constants. The definitions and assumptions of Section 3 must then be modified accordingly, e.g., Assumption B is modified to require that there exist $w_0 > 0$ such that the minimum eigenvalue of the matrix $\mathbf{w}(s)$ is not less than w_0 for every s.

Theorem 1 on existence holds with the following modifications:

(a) There exists a maximizer for which each component is a linear combination of an $(m-1)$-degree polynomial and the functions k_{T_i}.

(b) (i) This becomes the condition that there exists a maximizer of $L(\beta)$ with β restricted such that each component is an $(m-1)$-degree polynomial. (ii) This becomes the condition that (4.1) hold for all β and $\mathbf{h}$ whose components are $(m-1)$-degree polynomials, with $\mathbf{h}(s)^{\mathsf{T}}\mathbf{V}(\beta, s)\mathbf{h}(s)$ in the integrand.

There is essentially no change in the proof.

The argument relating to consistency of the estimator $\hat{\beta}$ proceeds along the lines of that for the one-dimensional case, with only modest modifications. The major novel points are extending Lemmas 5.1 and 5.2 of Silverman (1982) in a suitable way and modifying the definition of H_M to

$$\begin{aligned} H_M(\beta) = {} & \int_0^1\int_0^1\int_0^1 x[\beta(s) - \beta_0(s)]^{\mathsf{T}}\mathbf{V}_1(\beta_0(s) + ux[\beta(s) + \beta_0(s)]; s) \\ & [\beta(s) - \beta_0(s)]\,du\,dx\,d\bar{N}(s) \\ & - \frac{1}{n}\sum_{i=1}^{n}\int_0^1 [\mathbf{Z}_i - \mathbf{A}(\beta_0, s)]^{\mathsf{T}}[\beta(s) - \beta_0(s)]\,dN_i(s). \end{aligned}$$

Other modifications consist of changes in notation and minor manipulations.

The result is as follows.

THEOREM 6. *Suppose* $\alpha_n = O(n^{-\theta})$ *with* θ *satisfying* (6.21). *Then*

$$\max_{1 \le j \le p} \sup_t |\hat{\beta}_j(t) - \beta_{0j}(t)| \to_P 0 \tag{10.1}$$

at the rate given in (6.22).

A substantive difficulty arises in extending Theorem 3—on asymptotic normality—to the multivariate case: Although most of the argument goes through with only minor modifications, we have been unable to extend the inequality (7.7). Its derivation depends on an approximation to the eigenfunctions (ϕ_ν), as described in Naimark (1967). Although Naimark conjectures that the approximation extends to the multivariate case, we have not been able either to locate or to provide a proof.

Prentice–Self model. Prentice and Self (1983) generalize the ordinary Cox regression model by replacing the relative risk function $e^{\beta Z}$ by a more general form $r(\beta Z)$, where r is a function satisfying smoothness and positivity conditions; the motivating example [cf. (1.2)] is the linear function $r(y) = 1 + y$. They establish consistency and asymptotic normality of maximum partial likelihood estimators. Our model and results extend to this setting as well. Here we describe briefly what is involved.

The covariate-specific hazard function of (1.1) becomes

$$\lambda(t|\mathbf{z}) = \lambda_0(t) r\left(\sum_{j=1}^{p} \beta_{0j}(t) z_j \right), \tag{10.2}$$

where r is a function fulfilling assumptions stated below. The associated log-likelihood function [compare (2.1)] is thus

$$\begin{aligned} L(\beta) = (1/n) \sum_{i=1}^{n} D_i \left[\log r(\beta(T_i)Z_i) - \log\left(\sum_{j=1}^{n} Y_j(T_i) r(\beta(T_i)Z_j) \right) \right] \\ - \tfrac{1}{2}\alpha_n[\beta, \beta]. \end{aligned} \tag{10.3}$$

The function H of (3.1) then, of course, changes correspondingly. The definitions (Section 3) of the S_p and the s_p must be altered as in Prentice and Self (1983), changes in A, V, C, a, v and c are engendered as well. The function H_1 of (6.1) becomes

$$\begin{aligned} H_1(\beta) = \frac{1}{2}\alpha[\beta, \beta] + \frac{1}{2}\int_0^1 v(\beta_0, s) s_0(\beta_0, s) \lambda_0(s) [\beta(s) - \beta_0(s)]^2 \, ds \\ - \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left[Z_i \frac{r'(\beta_0(s)Z_i)}{r(\beta_0(s)Z_i)} - \frac{S_1(\beta_0, s)}{S_0(\beta_0, s)} \right] dN_i(s). \end{aligned} \tag{10.4}$$

In the definitions of the X_ν of (6.3), the function H_M of (6.8), the X_ν^* of (7.1) and the W_{ni} of (7.5), Z_i must be multiplied by $r'(\beta_0(s)Z_i)/r(\beta_0(s)Z_i)$ [the ratio is 1 for $r(y) = e^y$].

In addition to Assumptions A–D in Section 3, we require the following condition on the function r.

ASSUMPTION E. There exists $\varepsilon > 0$ such that

$$\inf_{\|\beta - \beta_0\|_\infty < \varepsilon} \; \inf_{s,\, z} r(\beta(s) z) > 0. \tag{10.5}$$

Given these definitional and notational changes, our results and their proofs carry over in the following manner.

1. While part (a) of Theorem 1 (regarding existence and computation of the MPPLE), whose proof does not depend on the form of r, remains valid as stated, neither (b) nor (c) seems to extend without senselessly restrictive conditions on r. It is possible, of course, that for specific choices of r, effective ad hoc computational methods may be developed.

2. Although (6.5) changes to reflect the altered definition of X_ν, the conclusions of Lemma 4, i.e., that $E[X_\nu] = 0$ and $\mathrm{Var}(X_\nu) \le 1/n$, hold as stated; the proof is unaltered.
3. Proposition 1 holds with the same proof.
4. Given Assumption E, Proposition 2 remains true, with no changes to the argument.
5. Proposition 3 and its antecedents continue to hold.
6. Consequently, Theorem 2 extends.
7. Lemmas 5 and 6, as well as Proposition 4, remain valid with no alterations (other than notational) to their proofs.
8. Thus, Theorem 3 holds as stated.
9. The additional results in Sections 8 and 9, therefore, also extend.

Time-dependent covariates. In many applications [cf. Andersen and Gill (1982)] it is desirable to allow covariates that vary over time. In our context this would correspond to a model in which the stochastic intensity λ_i for the counting process N_i is given by $\lambda_i(t) = \lambda_0(t)Y_i(t)e^{\beta(t)^T \mathbf{Z}_i(t)}$, where $\mathbf{Z}_i$ is a predictable stochastic process. Subject to conditions on $\mathbf{Z}$, our results extend with little change. The main restriction is that there must not exist a fixed time point t at which there is positive probability that $\mathbf{Z}$ will jump; this restriction is necessary in order to ensure continuity of $\mathbf{v}(\beta_0, t)$. Additional implicit restrictions on $\mathbf{Z}$ are engendered by differentiability conditions on w required for asymptotic normality. Also, moment conditions on $\mathbf{Z}$ are needed for Lemmas 1 and 2.

REFERENCES

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

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

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

BROWN, C. C. (1975). On the use of indicator variables in studying the time-dependence of parameters in a response-time model. *Biometrics* **31** 863–872.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. B* **34** 187–200.

DELVES, L. M. and MOHAMED, J. L. (1985). *Computational Methods for Integral Equations.* Cambridge Univ. Press, London.

GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 255–277.

GRAVES, L. M. (1927). Riemann integration and Taylor's theorem in general analysis. *Trans. Amer. Math. Soc.* **27** 163–177.

GRENANDER, U. (1981). *Abstract Inference.* Wiley-Interscience, New York.

HAHN, M. G. (1978). Central limit theorems in $D[0,1]$. *Z. Wahrsch. Verw. Gebiete* **44** 89–101.

KARR, A. F. (1986). *Point Processes and Their Statistical Inference.* Dekker, New York.

KARR, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model, via the method of sieves. *Ann. Statist.* **15** 473–489.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation to the partial sums of independent random variables and the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

McKEAGUE, I. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14** 579–589.

NAIMARK, M. A. (1967). *Differential Operators, Part I.* Ungar, New York.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. A* **135** 370–383.

O'SULLIVAN, F., YANDELL, B. S. and RAYNOR, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–103.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 803–814.

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

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.

TAPIA, R. A. and THOMPSON, J. W. (1978). *Nonparametric Probability Density Estimation.* The Johns Hopkins Univ. Press, Baltimore.

WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized smoothing spline problem. *Ann. Statist.* **13** 1378–1402.

NATIONAL HEART, LUNG
AND BLOOD INSTITUTE
BETHESDA, MARYLAND 20892

DEPARTMENT OF MATHEMATICAL SCIENCES
G. W. C. WHITING SCHOOL OF ENGINEERING
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218